PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/62, C07K 19/00, 14/505, C12N 15/85, 5/10, A61K 38/18 — A1

(11) International Publication Number: WO 99/02710

(43) International Publication Date: 21 January 1999 (21.01.99)

(21) International Application Number: PCT/US98/13944

(22) International Filing Date: 9 July 1998 (09.07.98)

(30) Priority Data:

08/890,929	10 July 1997 (10.07.97)	US
09/018,138	3 February 1998 (03.02.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	08/890,929 (CIP)
Filed on	10 July 1997 (10.07.97)
US	09/018,138 (CIP)
Filed on	3 February 1998 (03.02.98)

(71) Applicant *(for all designated States except US)*: BETH ISRAEL DEACONESS MEDICAL CENTER [US/US]; 330 Brookline Avenue, Boston, MA 02215 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: SYTKOWSKI, Arthur, J. [US/US]; 203 Park Avenue, Arlington, MA 02174 (US).

(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PRODUCTION AND USE OF RECOMBINANT PROTEIN MULTIMERS WITH ALTERED BIOLOGICAL ACTIVITY

(57) Abstract

The invention relates to recombinant fusion proteins, either with or without peptide linkers, with or without variations in noncoding regions of the protein. The fusion proteins have altered structural and functional properties and biological activity. The fusion proteins can be used, for example, as *in vivo* therapeutics.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PRODUCTION AND USE OF RECOMBINANT PROTEIN MULTIMERS WITH ALTERED BIOLOGICAL ACTIVITY

RELATED APPLICATIONS

This application is a continuation-in-part application of U.S. Serial No. 09/018,138, filed February 3, 1998, which is a continuation-in-part application of U.S. Serial No. 08/890,929, filed July 10, 1997, the teachings of which are incorporated herein by reference, in their entirety.

GOVERNMENT SUPPORT

This invention was made, in whole or in part, with Government support under Contract No. N00014-90-J-1847 awarded by the U.S. Navy. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

A problem encountered in the practice of medicine when using proteins as injectable pharmaceuticals is the frequency at which those injections must be made in order to maintain a therapeutic level of the protein in the circulation. For example, erythropoietin has a relatively short plasma half-life (Spivak, J.L., and Hogans, B.B., *Blood*, 73:90, 1989; McMahon, F.G., *et al.*, *Blood*, 76:1718, 1990). Therefore, therapeutic plasma levels are rapidly

decreased, and repeated intravenous administrations must be made. An alternative route of administration is subcutaneous injection. This route offers slower absorption from the site of administration, thus causing a sustained release effect. However, significantly lower plasma levels are achieved and, thus, a similar frequency of injection, as is required with intravenous administration, must be used to produce a comparable therapeutic effect.

Modification of naturally occurring proteins which have therapeutic value is often attempted in an effort to increase the protein's biological activity. Several methods have been employed to increase the biological activity of therapeutic proteins. These methods often focus on increasing the size of the therapeutic agents. For example, the size of a protein can be increased through chemical conjugation with a reagent such as polyethylene glycol (PEG) (Knusli, C., *et al.*, *Brit. J. Haematol.* 82:654-663, 1992). This procedure, also known as "PEGylation", has been reported with several protein agents, first as a means to reduce antigenicity, but also as a way to increase biological activity.

Another method of increasing a protein's size is through chemical cross-linking with another protein. For example, to increase the antigenicity of a protein, chemical cross-linking agents are used to conjugate the immunogenic protein to a carrier molecule such as immunoglobulin or serum albumin.

However, the conjugation of chemical compounds or inert molecules to a protein often results in a significant decrease of the overall biological activity, and of selected biological activity of the protein. (Knusli, C., *et al.*, *Brit. J. Haematol.*, 82:654-663, 1992). These conjugations must be designed such that the resulting modified protein remains therapeutically efficacious and retains the desired biological properties of the unmodified, wild type (i.e., naturally-occurring) protein (Satake, R., *et al.*, *Biochem. Biophys. Acta.* 1038:125-129, 1990). Thus, it would be advantageous to be able to modify therapeutically active proteins to increase their biological activity which would result in less frequent injections or smaller doses of protein.

SUMMARY OF THE INVENTION

The present invention relates to modified proteins or polypeptides with increased biological activity, and methods of producing and using these modified proteins and polypeptides.

Increased biological activity results from the production of fusion proteins that result in protein multimers, e.g., dimers and trimers. Protein multimers are produced by expressing tandemly linked nucleic acids encoding the proteins of the present invention, or biologically active fragments, analogs, variants, mutants or derivatives of the proteins. The nucleic acids that encode the proteins are fused, as described herein. The

proteins of the present invention can be fused directly to another protein, or can be fused via a linker, e.g., a peptide linker. The tandemly fused nucleic acid sequence is then inserted into an expression vector and introduced into a competent cell, either prokaryotic or eukaryotic, resulting in the production of a fusion protein multimer with increased biological activity.

Increased biological activity is defined herein as a prolonged plasma half-life (that is, a longer circulating half-life relative to the naturally occurring protein), or higher potency (i.e., requiring a smaller quantity relative to the naturally occurring protein to achieve a specified level of biological activity). Biological activity is also defined herein as greater affinity for receptors on target cells, or increased cell signaling (i.e., increased protein tyrosine kinase activity, induction of receptor clustering or aggregation, reduced receptor mediated endocytosis, reduced susceptibility to degradation such as proteolysis), or increased stability or half-life of the fusion protein during protein synthesis or the transcript encoding the fusion protein described herein. Increased biological activity can also encompass a combination of the above-described activities, for example, a modified protein with higher potency that also exhibits a prolonged circulating half-life. Because the proteins of the present invention have increased biological activity, the frequency with which they must be administered is reduced, or the amount administered to achieve an effective dose is reduced.

Additional advantages can also result from the modifications described herein. For example, new unpredictable activities can result, such as increased affinity for receptors or binding ligands, which can result in increased stimulation of signal generated by such binding. A reduced quantity of modified protein would then be necessary over the course of treatment as compared to the quantity necessary if unmodified protein were used.

Alternatively, or additionally, advantages can be augmented secretion rates and increased stability of the fusion protein during protein synthesis from host cells thereby increasing the overall yield of the fusion protein produced by, for example, *in vitro* methods or when introduced into a host cell for *in vivo* therapeutic purposes.

Proteins encompassed by the present invention include any protein with therapeutic activity. Specifically encompassed by the present invention are cytokines, growth factors, and hormones which include, for example, the following: Interferon-α, Interferon-β, Interferon-γ, Interleukin-1, Interleukin-2, Interleukin-3, Interleukin-4, Interleukin-5, Interleukin-6, Interleukin-7, Interleukin-8, Interleukin-9, Interleukin-10, Interleukin-11, Interleukin-12, Interleukin-13, Interleukin-14, Interleukin-15, Interleukin-16, Erythropoietin, Colony-Stimulating Factor-1, Granulocyte Colony-Stimulating Factor, Granulocyte-Macrophage Colony-Stimulating Factor, Leukemia Inhibitory Factor, Tumor Necrosis Factor, Lymphotoxin, Platelet-

Derived Growth Factor, Fibroblast Growth Factors, Vascular Endothelial Cell Growth Factor, Epidermal Growth Factor, Transforming Growth Factor-β, Transforming Growth Factor-α, Thrombopoietin, Stem Cell Factor, Oncostatin M, Amphiregulin, Mullerian-Inhibiting Substance, B-Cell Growth Factor, Macrophage Migration Inhibiting Factor, Endostatin, and Angiostatin. Exemplary descriptions and discussions of many of these proteins, including additional references for the manufacturing of proteins by recombinant technologies, protein purification schemes, and assessment of biological activity, can be found in "*Human Cytokines: Handbook for Basic and Clinical Research*", Aggarwal, B.B., and Gutterman, J.U., Eds., Blackwell Scientific Publications, Boston, MA, (1992), which is herein incorporated by reference in its entirety.

More specifically, the present invention relates to modified erythropoietin with increased biological activity, as defined herein. The modified erythropoietin with increased biological activity of the present invention is a fusion protein comprising two or more erythropoietin molecules covalently fused, resulting in an erythropoietin multimer.

Also encompassed by the present invention are methods of making and using the fusion protein multimers described herein and methods of using them.

The present invention provides fusion proteins with increased biological activity. As a result of the present

invention, the fusion proteins described herein provide proteins with improved therapeutic value.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram showing an EPO-EPO dimer DNA construct comprising an EPO A cDNA linked to an EPO B cDNA via a linker sequence.

Figure 2 is a diagram showing production of EPO A DNA and sequential elongation of linking DNA strand encoding $[gly\text{-}gly\text{-}gly\text{-}gly\text{-}ser]_3$ using PCR.

Figure 3 is a diagram showing production of EPO B DNA using PCR.

Figure 4 is a diagram showing the end product of each of steps II-IV of Figure 2.

Figure 5 is a diagram showing the end product of Figure 3.

Figure 6 is a diagram showing the restriction digest of pCRBlunt-EPO A^{IV} with Bgl I.

Figure 7 is a diagram showing the restriction digest of pCRBlunt-EPO B^{I} with Bgl I.

Figure 8 is a diagram showing the restriction digest of pCRBlunt-EPO $A^{IV}(-)$.

Figure 9 is a diagram showing the restriction digest of pCRBlunt-EPO $B^{I}(-)$.

Figure 10 is a diagram showing the Xho I and BamH I restriction digest of pcDNA3.1(-).

Figure 11 is a diagram showing the restriction digest of pcDNA3.1-EPO-EPO.

Figure 12 is a schematic representation discribing how differences in mRNA and protein structure; and protein function can result from alterations in the 5' and 3' UTR of a gene.

Figures 13 A-C depict the nucleotide sequence of the human EPO gene (SEQ ID NO: 1).

Figures 14 A-F depict the nucleic acid sequence of nucleotides 401-624 in the 5' untranslated region of the EPO gene (SEQ ID NO: 2) (Figure 14 A) and five variant sequences (SEQ ID NOS: 3-7) (Figures 14 B-F).

Figures 15 A-E depicts the nucleic acid sequence of nucleotides 2773-2972 in the 3' untranslated region of the EPO gene (SEQ ID NO: 8) (Figure 15 A) and four variant sequences (SEQ ID NOS: 9-12) (Figures 15 B-E).

Figures 16A, 16B and 16C depict the nucleic acid sequence (SEQ ID NO: 16) and corresponding amino acid sequence (SEQ ID NO: 17) of the erythropoietin dimer EPO^{wt}-L-EPO^{wt}. A polypeptide linker of 17 amino acid in length (L) links the two EPO protein molecules.

Figures 17A, 17B and 17C are graphic representations of the *in vivo* efficacy of the erythropoietin dimer fusion protein (EPO-EPO), erythropoietin monomer (EPO) and culture media from nontransfected COS1 cells (Control) as measured by changes in hematocrits obtained before (Pre) and 7 days after (Post) the administration of a single dose of 300 IU/kg protein.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the term fusion protein refers to the fusion of one protein molecule with another protein molecule. In one embodiment the C-terminus of one protein molecule is fused to the N-terminus of another protein molecule. In another embodiment the N-terminus of one protein is linked to the C-terminus of another protein molecule. The fusion proteins of the present invention typically include constructs in which a linker peptide sequence is utilized. The fusion proteins of the present invention have a formula of R_1-R_2 or R_1-L-R_2, where R_1 and R_2 are substantially similar or identical protein molecules and L is a linker, typically a peptide. In another embodiment of the present invention, R_1 and R_2 can be different proteins. R1 and R2 are also referred to as the monomeric subunits of the fusion proteins described herein. The protein molecules are fused to one another in such a manner as to produce a single fusion protein comprised of two or more protein molecules. The fusion protein produced has increased biological activity. In one embodiment of the present invention the protein molecules are EPO.

Fusion protein constructs are named by listing the respective molecules. For example, EPO-L-EPO refers to a fusion protein comprised of two EPO molecules joined by a peptide linker, and EPO-L-EPO-L-EPO refers to a fusion protein comprised of three EPO molecules joined by two peptide linkers.

The invention also relates to isolated nucleic acid constructs which encode, either independently or together, the protein molecules of the fusion proteins of the present invention. Nucleic acid constructs are defined herein as heteropolymers of nucleic acid sequences. Nucleic acid sequences are meant to refer to chains of nucleotides joined together by phosphodiester bonds to form a nucleic acid heteropolymer. The nucleic acid sequences can be double stranded or single stranded. The nucleic acid sequence can include one or more exons, with or without, as appropriate, introns. Methods to generate nucleic acid constructs encoding the fusion proteins of the invention are standard molecular biological procedures and well known in the art. (See, for example, Aubusel, *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons, Inc (1997). Alternative combinations or modifications of the nucleic acid constructs according to the present invention would be apparent to the person of skill in the art.

Specifically encompassed are nucleic acid constructs encoding fusion proteins of EPO. For example, a nucleic acid construct encodes an EPO dimer (EPO-L-EPO; Figures 16A-16C SEQ ID NOS: 16 and 17) wherein the nucleic acid encoding the amino terminal portion of the EPO dimer (R1) comprises the 5' untranslated region, a start codon, a pre-protein including the leader sequence, yet lacks a stop codon and is followed by a 17 amino acid peptide linker. The nucleic acid construct encoding the carboxy terminal portion of the EPO dimer (R2) comprises the coding region

for the mature secreted EPO, a stop codon, and the 3' untranslated region. When expressed in COS1 cells the mRNA encoding the EPO-L-EPO is 2.8 kb in length and the secreted protein dimer is 76 kDa, a molecular weight approximately equivalent to two fully processed and glycosylated EPO protein molecules (37 kDa each) and the 17 amino acid peptide linker (1.8 kDa) (See Example 2). As used herein EPO-L-EPO, EPO-EPO and EPO^{wt}-EPO^{wt} are used equivalently to refer to erythropoietin dimers wherein the monomeric EPO is the wildtype EPO (EPO^{wt}).

The EPO^{wt}-L-EPO^{wt} dimer has increased biological activity compared to monomeric EPO^{wt} when assessed by the standard, art-recognized *in vitro* bioassay of Krystal (*Exp. Hematol. 11*:649-660 (1983)) and by the *in vivo* ability to augment hematocrits following multiple or single injections of dimer (See Example 5; Figures 17A-17C). The EPO^{wt}-EPO^{wt} protein dimer was secreted at significantly higher amounts than EPO^{wt} monomer. Epo^{wt} was secreted from COS1 cells at levels of 6.3 U/ml (0.018 µg/ml), and exhibited a specific activity of 350 U/µg. However, Epo^{wt}/Epo^{wt} was secreted at a level of 151 U/ml (0.150 µg/ml) and had a specific activity of 1007 U/µg, approximately three fold higher than monomeric EPO (See Examples 4 and 5).

Subcutaneous injection of a single dose of EPO^{wt}-EPO^{wt} (300 U/kg) to mice resulted in an elevated hematocrit seven days later, (Example 5, Figures 17A-17C). These *in vivo* and *in vitro* data clearly show the unexpected nature of increased biological activity and enhanced secretion of

dimeric EPO molecules which can have important implications for *in vitro* production and *in vivo* pharmaceutical compositions.

The nucleic acid constructs of the invention can encode the wildtype protein (e.g., EPO^{wt}) or mutant EPO (e.g., wherein a mutation occurs in the coding region of at least one of the monomeric subunits). The mutation can be, for example, the R103A mutation in EPO wherein the arginine at position 103 is replaced with an alanine. Specifically encompassed by the present invention are EPO^{R103A}-EPO^{wt}, EPO^{wt}-EPO^{R103A}, and EPO^{R103A}-EPO^{R103A} mutant dimers (Examples 7 and 8). A single point mutation in the coding region of EPO results in a dimer with increased biological activity compared to wildtype monomeric EPO. The secretion of the mutant dimers EPO^{R103A}-EPO^{wt}, EPO^{wt}-EPO^{R103A}, and EPO^{R103A}-EPO^{R103A} was unexpectedly enhanced compared to the EPO^{wt}-EPO^{wt} dimer suggesting that the R103A mutation can increase stability of EPO dimer proteins.

The fusion proteins of the present invention can be produced by recombinant DNA technologies using nucleic acid constructs (Examples 1 and 7). The term "recombinant", as used herein, means that a protein is derived from recombinant (e.g., eukaryotic or prokaryotic host cell) expression systems which include, for example, yeast (e.g., Saccharomyces), bacteria (such as, Escherichia or Bacillus), and animal cells including insect or mammalian expression systems. Proteins expressed in most bacterial cultures will be free of glycan. Protein expressed in

yeast may have a glycosylation pattern different from protein expressed in mammalian cells.

As used herein, the term nucleotide sequence or nucleic acid sequence refers to a heteropolymer of deoxyribonucleotides (DNA), or ribonucleotides (RNA).

Nucleic acid sequences encoding the proteins provided in this invention can be assembled from DNA, either cDNA or genomic DNA, or RNA, and short oligonucleotide linkers to provide a synthetic nucleic acid sequence which is capable of being expressed in a recombinant transcriptional unit.

Homologous nucleic acids, including DNA or RNA, can be detected and/or isolated by hybridization (e.g., under high stringency conditions or moderate stringency conditions). "Stringency conditions" for hybridization is a term of art which refers to the conditions of temperature and buffer concentration which permit hybridization of a particular nucleic acid to a second nucleic acid in which the first nucleic acid may be perfectly complementary to the second, or the first and second may share some degree of complementarity which is less than perfect. For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions" and "moderate stringency conditions" for nucleic acid hybridizations are explained in several technical protocol reference texts, for example, Ausubel, F.M., *et al.*, "*Current Protocols in Molecular Biology*" (1995), the teachings of which are hereby incorporated by

reference. The exact conditions which determine the stringency of hybridization depend not only on ionic strength, temperature and the concentration of destabilizing agents such as formamide, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, high or moderate stringency conditions could be determined for detecting the various forms of recombinant polypeptides.

By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (e.g., selectively) with the sequences, with substantially similar identity in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and Aaronson, S.A., *Methods in Enzymology, 200*:546-556, 1991. Also, *"Current Protocols in Molecular Biology"* (*supra*), which describes how to determine washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that

hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined for high, moderate or low stringency, depending on the level of mismatch sought. For example, in this invention alterations in the noncoding (5' and 3' untranslated) regions of the gene may necessitate changes in stringency conditions from low to medium to high depending upon the number of nucleotides that are modified that differ from the condition used to detect wild type versions of the gene. Where appropriate the salt concentrations and temperatures will be adjusted accordingly.

The term recombinant expression vector, as used herein, refers to a replicable DNA construct used either to amplify or to express DNA which encodes the fusion proteins of the present invention. The recombinant expression vector includes a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein; and (3) appropriate transcription and translation initiation and termination sequences. Structural elements intended for use in yeast expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader sequence or transport sequence, it may include an N-terminal methionine residue.

This residue may optionally be subsequently cleaved from the expressed recombinant protein to provide a final product.

A DNA sequence encoding a fusion protein is constructed using recombinant DNA techniques to assemble separate DNA fragments encoding proteins into an appropriate expression vector. For example, the 3' end of a DNA molecule encoding a protein is ligated to the 5' end of a second DNA molecule encoding the same or a substantially similar protein, with the reading frames of the sequences in phase to permit mRNA translation of the sequences into a single biologically active fusion protein. The DNA molecules are joined in tandem, meaning that the DNA molecules are joined in succession, one after the other. The regulatory elements responsible for transcription of DNA into mRNA are retained on the first of the two DNA sequences while binding signals or stop codons, which would prevent read-through to the second DNA sequence, are eliminated. Conversely, regulatory elements are removed from the second DNA sequence while stop codons required to end translation are retained.

As described herein, means are provided for linking protein molecules, preferably via a linker sequence(s). The linker sequence(s) separates the protein molecules by a distance sufficient to ensure that each protein molecule properly folds into its secondary and tertiary structures. Suitable linker sequences (1) adopt a conformation suitable to result in a fusion protein with increased biological

activity, (2) do not exhibit a propensity for developing an ordered secondary structure which could impair the biological functions of the protein molecules, and (3) have minimal hydrophobic or charged character which could impair the biological functions of the EPO molecules. For example, a suitable linker will produce a fusion protein where interaction of the protein components results in increased biological activity. The linker conformation can be flexible or rigid, depending on the final conformation of the fusion required to result in increased biological activity. An example of a more rigid linker would be a linker with an α-helix that would not allow free rotation of the linked protein components. Typical surface amino acids in flexible protein regions include Glycine (Gly), Asparagine (Asn) and Serine (Ser). Virtually any permutation of amino acid sequences containing Glycine (Gly), Asparagine (Asn) and Serine (Ser) would be expected to satisfy the above criteria for linker sequence. Other near neutral amino acids, such as Threonine (Thr) and Alanine (Ala), may also be used in the linker sequence.

The length of the linker sequence may vary without significantly affecting the biological activity of the fusion protein. Generally, the protein, e.g., EPO, molecules will be separated by a linker sequence having a length of about 10 amino acids to about 20 amino acids, although longer linker sequences may be used, for example, a full-length polypeptide can comprise the linker. In the most preferred aspects of the present invention, the linker

sequence is about 15 amino acids in length. The linker sequence is incorporated into the fusion protein by well-known methods and as described in Example 1.

In one embodiment, the fusion proteins described herein comprise wild type (e.g., naturally-occurring) proteins with therapeutic activity. As defined herein, therapeutic activity means the ability of a fusion protein, upon administration to a mammal, to alleviate, to any degree, or eliminate the deficiency or condition for which the mammal is being treated. Specifically encompassed by the present invention are cytokines, growth factors, and hormones which include, for example, the particular proteins listed in the following paragraphs followed by the appropriate reference(s). These references as described herein, provide guidance for the production, purification and evaluation of biological activity of the cytokines using standard, routine methods. Nucleic acid sequences coding for the protein molecules of suitable cytokines (e.g., EPO) for making the protein multimers of the present invention are known in the art and can be readily obtained from, for example, the EMBL/GenBank data bases. Each of the references in the following paragraphs is incorporated by reference in its entirety.

INTERFERON-α: Henco, K., *et al.*, *J. Mol. Biol.*, 185: 227-260 (1985). Pestka, S., *et al.*, *Ann. Rev. Biochem.*, 56: 727-777 (1987). *Methods in Enzymology*, Pestka, S., (Ed.), Academic Press, New York, NY, 119:3-14 (1986).

INTERFERON-β: "*Human Cytokines: Handbook for Basic and Clinical Research*", Aggarwal, B.B., and Gutterman, J.U. (Eds.), Blackwell Scientific Publications, Boston, MA (1992).

INTERFERON-γ: Gray, P.W., *et al.*, *Nature*, 298:859-863 (1982). Rinderknecht, E., *et al.*, *J. Biol. Chem.*, 259:6790-6797 (1984).

INTERLEUKIN-1: IL-1α: Furutani, Y., *et al.*, *Nucleic Acids Res.*, 143:167-3179 (1986). IL-1β: Clark, B.D., *et al.*, *Nucleic Acids Res.* 14:7897-7914 (1986).

INTERLEUKIN-2: Fujita *et al.*, 1983. Durand, D.B., *et al.*, *Mol. Cell Biol.*, 8:1715-1724 (1988).

INTERLEUKIN-3: Yang, Y.C., *et al.*, *Cell*, 47:3-10 (1986).

INTERLEUKIN-4: Arai, N., *et al.*, *J. Immunol.*, 142:274-282 (1989).

INTERLEUKIN-5: Azuma, C., *et al.*, *Nucleic Acids Res.*, 14:9149-9158 (1986). Yokota, T., *et al.*, *Proc. Natl. Acad. Sci.* USA, 84:7388-7392 (1987).

INTERLEUKIN-6: Hirano, T., *et al.*, *Nature*, 324:73-76 (1986). Van Snick, J., *et al.*, *Eur. J. Immunol.*, 18:193-197 (1988).

INTERLEUKIN-7: Goodwin, R.G., *et al.*, *Proc. Natl. Acad. Sci.* USA, 86:302-306 (1989).

INTERLEUKIN-8: Kusner, D.J., *et al.*, *Kidney International* 39:1240-1248 (1991).

INTERLEUKIN-9: Renauld, J-C., *et al.*, *J. Immunol.*, 144:4235-4241 (1990). Moeller, J., *et al.*, *J. Immunol.*

144:4231-4234 (1990). Yang, Y.C., *et al.*, *Blood*, 74:1880-1884 (1989).

INTERLEUKIN-10: Moore, K.W., *et al.*, *Science*, 248:1230-1234 (1990). Fiorentino, D.F., *et al.*, *J. Exp. Med.*, 170:2081-2095 (1989).

INTERLEUKIN-11: Paul, S.R., *et al.*, *Proc. Natl. Acad. Sci.* USA, 87:7512-7516 (1990).

INTERLEUKIN-12: Wolf, S.F., *et al.*, *J. Immunol.*, 146:3074-3081 (1991); BLAST Database (www.ncib.nlm.nih.gov), accession number M65290.

INTERLEUKIN-13: Dolganov, G., *Blood*, 87:3316-3326 (1996).

INTERLEUKIN-14: Ambrus, J.L., *et al.*, *Proc. Natl. Acad. Sci.* USA, 90:6330-6334 (1993).

INTERLEUKIN-15: Meazza, R., *et al.*, *Oncogene*, 12:2187-2192 (1996).

INTERLEUKIN-16: Cruikshank, W.W., *et al.*, *Proc. Natl. Acad. Sci.* USA, 91:5109-5113 (1994).

ERYTHROPOIETIN: Jacobs, K., *et al.*, *Nature*, 313:806-810 (1985).

COLONY-STIMULATING FACTOR-1: Kawasaki, E.S., *et al.*, *Science*, 230:291-296 (1985). Wong, G.G., *et al.*, *Science*, 235:1504-1508 (1987). Ladner, M.B., *et al.*, *EMBO. J.*, 6:2693-2698 (1987). Cerretti, D.P., *et al.*, *Mol. Immunol.*, 25:761-770 (1988). "*Colony Stimulating Factors*", Dexter, T.M., *et al.* (Eds.), Marcel Dekker Publishers, New York, NY pp. 155-176 (1990).

GRANULOCYTE-COLONY-STIMULATING FACTOR: Nagata, S., *et al.*, *Nature*, 319:415-418 (1986). Souza, L.M., *et al.*, *Science*, 232:61-65 (1986).

GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR: Miyataka, S., *et al.*, *EMBO J.*, 4:2561-2568 (1985).

LEUKEMIA INHIBITORY FACTOR: Moureau, J-F., *et al.*, *Nature*, 336:690-692 (1988).

TUMOR NECROSIS FACTOR: Nedwin, G.E., *et al.*, *Nucleic Acids Res.*, 13:6361-6373 (1985).

LYMPHOTOXIN: Nedwin, G.E., *et al.*, *J. Cell Biochem.*, 29:171-182 (1985).

PLATELET-DERIVED GROWTH FACTOR: Deuel, T.F., *et al.*, *J. Biol. Chem.*, 256:8896-8899 (1981). "*Human Cytokines: Handbook for Basic and Clinical Research*", Aggarwal, B.B., and Gutterman, J.U. (Eds.), Blackwell Scientific Publications, Boston, MA (1992).

FIBROBLAST GROWTH FACTORS: Abraham, J.A., *et al.*, *Science*, 233:545-547 (1986a).

VASCULAR ENDOTHELIAL CELL GROWTH FACTOR: Keck, P.J., *et al.*, *Science*, 246:1309-1312 (1989).

EPIDERMAL GROWTH FACTOR: Scott, J., *et al.*, *Science*, 221:236-240 (1983). Gray, A., *et al.*, *Nature*, 303:722-725 (1983).

TRANSFORMING GROWTH FACTOR-β: Derynck, R., *et al.*, *Nature*, 316:701-705 (1985). Scotto, L., *et al.*, *J. Biol. Chem.*, 265:2203-2208 (1990).

TRANSFORMING GROWTH FACTOR-α: Derynck, R., *Cell*, 54:593-595 (1988).

THROMBOPOIETIN: Sohma, Y., *et al.*, *FEBS Lett.*, 353: 57-61 (1994); BLAST Database (www.ncib.nlm.nih.gov), accession number D32046.

STEM CELL FACTOR: Williams, D.E., *et al.*, *Cell*, 63:167-174 (1990). Copeland, N.G., *et al.*, *Cell*, 63:174-183 (1990). Flanagan, J.G., *et al.*, *Cell*, 63:185-194 (1990). Zsebo, K.M., *et al.*, *Cell*, 63:213-224 (1990). Martin, F.H., *et al.*, *Cell*, 63:203-211 (1990). Zsebo, K.M., *et al.*, *Cell*, 63:195-201 (1990). Huang, E., *et. al.*, *Cell*, 63:225-233 (1990). Anderson, D.M., *et al.*, *Cell*, 63:235-243 (1990).

ONCOSTATIN M: Linsley, P.S., *et al.*, *Mol. Cell. Biol.*, 10:1882-1890 (1990). Zarling, J.M., *et al.*, *Proc. Natl. Acad. Sci.* USA, 83:9739-9743 (1986). Malik, N., *et al.*, *Mol. Cell. Biol.*, 9:2847-2853 (1989).

AMPHIREGULIN: Plowman, G.D., *et al.*, *Mol. Cell. Biol.*, 10:1969-1981 (1990). Shoyab, M., *et al.*, *Proc. Natl. Acad. Sci.* USA, 85:6528-6532 (1988).

MULLERIAN-INHIBITING SUBSTANCE: Cate, R.L., *et al.*, *Cell*, 45:685-698 (1986). Wallen, J.W., *et al.*, *Cancer Res.*, 49:2005-2011 (1989). Picard, J-Y., *et al.*, *Proc. Natl. Acad. Sci.* USA, 83:5464-5468 (1986). Coughlin, J.P., *et al.*, *Mol. Cell. Endocrinol.*, 49:75-86 (1987).

B-CELL GROWTH FACTOR: Sharma, S., *et al.*, *Science*, 235:1489-1492 (1987).

MACROPHAGE MIGRATION INHIBITORY FACTOR: Weiser, W.Y., *et al.*, *Proc. Natl. Acad. Sci.* USA, 86:7522-7526 (1989).

ENDOSTATIN: O'Reilly, M.S., *et al.*, *Cell*, 88:277-285 (1997).

ANGIOSTATIN: O'Reilly, M.S., *et al.*, *Cell*, 79:315-328 (1994).

Also encompassed by the present invention are fusion proteins comprising biologically active fragments, analogs, mutants, variants or derivatives of the naturally-occurring proteins described herein. Biologically active fragments, derivatives, analogs, variants and mutants of the naturally-occurring proteins are also referred to herein as substantially similar proteins of the naturally-occurring protein. However, the level of biological activity of fragments, analogs, mutants, variants or derivatives of the naturally-occurring protein need not be identical to the activity of the naturally-occurring protein (also referred to herein as the parent protein). For example, a fragment of a cytokine protein may exhibit only 50-80% of the activity of the naturally-occurring cytokine, yet because two or more cytokines, either the same or different, are linked to form a fusion protein, the fusion protein exhibits increased biological activity as compared to a monomer of the naturally-occurring cytokine. Tests to determine biological activity are well-known to those of skill in the art and can include, for example, measuring the extent of hematopoiesis, platelet production, receptor binding, angiogenesis, immunostimulation, or immunosuppression. For example, the biological activity of a mutant of erythropoietin can be measured using *in vitro*

and *in vivo* assays as described in U.S. Patent Nos. 5,614,184 and 5,580,853. The teachings of which are herein incorporated by reference in their entirety.

As described in detail in Sytkowski et al., (U.S. Patent No. 5,614,184 (1997)) and in Example 4, the biological activity of EPO fusion proteins can be determined using the well-established *in vitro* bioassay of Krystal (*Exp. Hematol. 11*:649-660 (1983). The Krystal assay measures the effect of EPO on intact mouse spleen cells. EPO fusion protein-stimulated production of red blood cells is then monitored by ^{3}H-thymidine incorporation and quantitated by competitive radioimmunoassay (RIA) or enzyme linked immunoassay (ELISA). The specific activity of the EPO fusion protein is expressed as a ratio of international units according to the World Health Organization Second International Reference Preparation divided by micrograms of protein determined by RIA or ELISA. As a control the *in vitro* biological activity of EPO monomer is run in parallel with EPO fusion proteins. EPO-EPO dimers had approximately eight-fold higher biological activity compared to monomeric protein.

The biological activity of EPO fusion proteins can also be evaluated using *in vivo* assays according to the method of Sytkowski A.T., *et al., (Proc. Natl. Acad. Sci. USA 95:*1184-1188 (1998)) as described in Example 5. Briefly, EPO multimer (e.g., 300 IU/kg) is injected into mice and hematocrits determined in blood samples obtained before (Pre) or after (Post) treatment. EPO multimer can

be administered on days 1, 3 and 5; and hematocrits determined on day 8. Additionally, or alternatively, mice can receive a single injection of EPO multimer on day 1 and the hematocrit determined 7 days following the injection or on day 8 of the assay.

As shown in Figures 17A-17C and in Example 5 a single injection of 300 IU/kg of the dimer EPO^{wt}-L-EPO^{wt} results in an increase in the mean hematocrit compared to animals injected with wildtype monomer. Thus, the hematocrit of EPO-EPO dimer treated mice remained elevated at day 8 unlike the monomer treated animals. Therefore, the half-life and *in vivo* activity of dimerized erythropoietin was augmented. These *in vivo* data for fusion proteins described herein are significant in documenting biologically potent fusion proteins with enhanced activity and prolonged half-lives. The observed increase in *in vitro* and *in vivo* bioactivity of EPO-EPO dimers are, unexpectedly significantly greater than activities that would be predicted from two molecules of EPO. Indeed, less frequent, for example, subcutaneous administration of polypeptides in a clinical setting can be therapeutically efficacious.

The present invention also provides fusion proteins with or without associated native-protein glycosylation. Non-glycosylated fusion proteins can be expressed from nucleic acid constructs in host cells, such as *E. coli*, which do not glycosylate protein molecules. Alternatively, or additionally, nucleic acid constructs encoding the

fusion proteins of the present invention can be selectively modified to encode mutant analogs having inactivated N-glycosylation sites by routine methods such as oligonucleotide synthesis and ligation or site-specific mutagenesis (See Example 6). These analog proteins can be produced in a homogeneous, reduced carbohydrate form in good yield using yeast expression systems. N-glycosylation sites in eukaryotic proteins are characterized by the amino acid triplet Asn-A_1-Z, where A_1 is any amino acid except Proline (Pro), and Z is Serine (Ser) or Threonine (Thr). In this sequence, asparagine provides a side chain amino group for covalent attachment of carbohydrate. Such a site can be eliminated by substituting another amino acid for Asparagine (Asn) or for residue Z, deleting Asparagine (Asn) or Z, or inserting a non-Z amino acid between A_1 and Z, or an amino acid other than Asparagine (Asn) between Asparagine (Asn) and A_1.

Also encompassed by the present invention are fusion proteins comprising protein molecules with at least one or more additional glycosylation sites (e.g., N-linked, or O-linked) than the wildtype protein molecule. The protein molecules can be, for example, novel erythropoiesis stimulating protein molecules, such as NESP, which has an increased circulatory half-life (Egrie, J., *et al.*, *Blood* *90*:56a (1997); Fürst, I., *Nature Biotechnology* *15*:940 (1997)). In the case of NESP, additional N-linked sites can be generated as described herein by, for example, substitution of leucine at position 69, or alanine at

position 125 (or both) with an asparagine residue; or, additionally or alternatively, alanine at position 127 with a serine residue. Likewise, additional O-linked glycosylation sites can be produced wherein the alanine at position 123 or position 125 is substituted with a threonine or serine residue. Glycosylation can be further enhanced by placing proline residues at positions -1 and/or +1 relative to the glycosylation sites (e.g., the glycosylation of a threonine residue at position 125 is augmented when proline is located at either position 124 or position 126) (Elliott, S., *et al.*, *Biochemistry 33*:11237 (1994)). It is further envisioned that any combination of the above described glycosylation sites would also be within the scope of the invention.

Derivatives and analogs can be obtained by mutations of the fusion protein. A derivative or analog, as referred to herein, is a polypeptide comprising an amino acid sequence that shares sequence identity, or similarity, with the full-length sequence of the wild type (or naturally occurring protein), except that the derivative or analog has one or more amino acid sequence differences attributable to a deletion, insertion and/or substitution. Sequences for derivatives or analogs of the protein molecules comprising the fusion proteins of the present invention can be aligned using database search strategies well known in the art including, for example, Basic Local Alignment Search Tool (BLAST) (Altschul, S.F., *et al.*, *J. Mol. Biol. 215*:403-410 (1990)) and FASTA (Pearson, W.R., *et*

al., Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448 (1988)) algorithms.

Bioequivalent analogs of proteins can be constructed by, for example, making various substitutions of residues or sequences. For example, cysteine residues can be deleted or replaced with other amino acids to prevent formation of incorrect intramolecular disulfide bridges upon renaturation. Other approaches to mutagenesis involve modification of adjacent dibasic amino acid residues to enhance expression in yeast systems in which KEX2 protease activity is present. Generally, substitutions should be made conservatively; i.e., the most preferred substitute amino acids are those having physicochemical characteristics resembling those of the residue to be replaced. Similarly, when a deletion or insertion strategy is adopted, the potential effect of the deletion or insertion on biological activity should be considered. Due to degeneracy of the genetic code, there can be considerable variation in nucleotide sequences encoding the same amino acid sequence. Therefore, the nucleic acid constructs encoding the fusion proteins of the present invention, bioequivalent analogs or derivatives of the protein molecules comprising the fusion proteins described herein can be altered to produce a codon encoding the same amino acid. For example, the amino acid alanine is encoded by the nucleotide triplet GCA or GCC or GCG or GCU. This can be advantageous for the production of fusion proteins

by recombinant technologies where a codon is preferred by a selected host cell.

Mutations in nucleotide sequences constructed for expression of analogs must, of course, preserve the reading frame phase of the coding sequences and preferably will not create complementary regions that could hybridize to produce secondary mRNA structures such as loops or hairpins which would adversely affect translation of the mRNA. Alternatively, mutations could introduce secondary structure which would result in higher translational efficiency. Although a mutation site may be predetermined, it is not necessary that the nature of the mutation *per se* be predetermined. For example, in order to select for optimum characteristics of mutants at a given site, random mutagenesis may be conducted at the target codon and the expressed mutants screened for the desired activity.

Mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion, substitution, or deletion. Alternatively, oligonucleotide-directed site-specific mutagenesis procedures (see Example 6) can be employed to provide an altered gene having particular codons altered according to the substitution, deletion, or insertion required. Conserved and/or nonconserved amino acids, including amino acids at or distant from the active site of

the protein molecules comprising the fusion proteins, can be modified. Exemplary methods of making the alterations set forth herein are disclosed by Walder *et al.* (*Gene* 42:133, 1986); Bauer *et al.* (*Gene* 37:73, 1985); Craik (*BioTechniques*, January 1985, 12-19); Smith *et al.* ("*Genetic Engineering: Principles and Methods*", Plenum Press, 1981); and U.S. Patent Nos. 4,518,584 and 4,737,462 and are incorporated by reference herein. Such techniques are routine, art-recognized and well-known to one of skill. Generally, commercially available kits to make the alterations described herein are available.

The mutations can be in one or more of the protein molecules comprising the fusion protein. Specifically, the present invention encompasses mutants of EPO fusion proteins comprising a R103A mutation in the coding region of EPO wherein the arginine at position 103 is replaced by an alanine in at least one of the EPO domains of the dimer (See Example 7). Such a mutation can lead to increased stability of the mRNA encoding for the mutant dimer fusion protein thereby resulting in increased secretion of the fusion protein providing a useful method for augmenting production yields of fusion proteins using recombinant methods described herein as well as for *in vivo* therapeutics. Other mutations in the EPO protein molecules comprising the fusion protens of the present invention can also be made. For example, the substitution of the glycine residue at position 101 with another amino acid such as alanine, which results in increased biological activity of

EPO monomer (Sytkowski *et al.*, U.S. Patent No. 5,614,184 (1997)), could be introduced into EPO protein molecules.

Other EPO mutants which comprise the protein molecules of the fusion proteins of the present invention can be, for example, produced as described by Elliott *et al.*, (*Biochemistry* 33:11237 (1994)). For example, Elliott *et al.*, have shown that biological activity is not altered by site specific mutagenesis within residues 21-44; 52-95; 109-140; and 163-166 of EPO. Thus, fusion proteins comprising EPO mutant protein molecules with these mutations are within the scope of the invention.

Peptidomimetics (molecules which are not protein molecules, but which mimic aspects of their structures to mediate biological activity), that are based upon the fusion proteins (e.g., EPO^{wt}-L-EPO^{wt}, EPO^{wt}-L-EPO^{R103A}, EPO^{R103A}-EPO^{wt}, EPO^{R103A}-EPO^{R103A}) and protein molecules (e.g., EPO) comprising the fusion proteins of the present invention, which have increased biological activity(e.g., increased hematocrit, prolonged protein or mRNA half-life), are also within the scope of the present invention. For example, polysaccharides can be prepared that have the same functional groups as the fusion proteins and protein molecules of the invention, and which interact with target cells or mediate biological activity in a similar manner. Peptidomimetics can be designed, for example, by establishing the three dimensional structure of a protein molecule in the environment in which it is bound or will bind to the target cell.

The protein molecules of the present invention can comprise naturally-occurring amino acids (e.g., L-amino acids), non-naturally-occurring amino acids (e.g., D-amino acids), and small molecules that biologically mimic the protein molecules such as the protein molecule analogs, derivatives or mimetics described herein. The protein molecules and peptidomimetics of the present invention can be in a linear or cyclic conformation.

The protein molecules and fusion proteins of the present invention can comprise either the 20 naturally occurring amino acids or other synthetic amino acids. Synthetic amino acids encompassed by the present invention include, for example, naphthylalanine, L-hydroxypropylglycine, L-3,4-dihydroxyphenylalanyl, α-amino acids such as L-α-hydroxylysyl and D-α-methylalanyl, L-α-methyl-alanyl, ß amino-acids such as ß-analine, and isoquinolyl.

D-amino acids and other non-naturally occurring synthetic amino acids can also be incorporated into the protein molecules and fusion proteins of the present invention. Such other non-naturally occurring synthetic amino acids include those where the naturally occurring side chains of the 20 genetically encoded amino acids (or any L or D amino acid) are replaced with other side chains, for instance with groups such as alkyl, lower alkyl, cyclic alkyl, amide, hydroxy, carboxy, lower alkyl carboxylic acid ester, sulfonic acid, a lower alkyl sulfonic acid ester or a phosphorous acid or ester thereof.

These peptidomimetics possess biological activity (e.g., stimulate hematopoiesis) lesser than, comparable to or greater than the biological activity of the corresponding protein molecule (e.g., EPO) or fusion protein (e.g., EPO^{wt}-L-EPO^{wt}, EPO^{wt}-L-EPO^{R103A}, EPO^{R103A}-EPO^{wt}, EPO^{R103A}-EPO^{R103A}), but can possess a "biological advantage" over the corresponding protein with respect to one, or more, of the following properties: solubility, stability, and susceptibility to hydrolysis or proteolysis.

Methods for preparing peptidomimetics include modifying the N-terminal amino group, the C-terminal carboxyl group, and/or changing one or more of the amino linkages in the peptide to a non-amino linkage. Modifications of peptides to produce peptide mimetics are described in U.S. Patent Nos: 5,643,873 and 5,654,276, the teachings of which are incorporated herein by reference. The protein molecules and fusion proteins can also be cyclic peptide mimetics. Such cyclic test substances can be produced using known laboratory techniques (e.g., as described in U.S. Patent No: 5,654,276, the teachings of which are herein incorporated in their entirety by reference).

These peptidomimetic compounds can be manufactured by art-known and art-recognized methods. Determining an appropriate chemical synthesis route for the peptidomimetics will depend upon the particular protein molecule and fusion protein and is generally readily identified using no more than routine skill.

For example, when designing a peptidomimetic for a basic amino acid in a protein molecule suitable nitrogen containing groups can include amines, ammoniums, guanidines and amides or phosphoniums; or when designing a peptidomimetic for an acidic amino acid in a protein molecule, a carboxyl, lower alkyl carboxylic acid ester, sulfonic acid, a lower alkyl sulfonic acid ester or a phosphorous acid or ester thereof can be used. The nitrogen of a peptide bond can be substituted with oxygen or sulfur, thereby forming a polyester backbone. Likewise, the carbonyl of the peptide bond can be substituted with a sulfonyl group or sulfonyl group, thereby forming a polyamide. Reverse amides of protein molecules can also be made (e.g., substituting one or more -CONH- groups for a -NHCO- group). In addition, the peptide backbone can be substituted with a polysilane backbone.

In another embodiment, the fusion proteins described herein comprise variant type proteins produced by modifications in 5' and/or 3' noncoding regions of the wildtype gene. Hereinafter, the term recombinant variant protein will be used to describe these molecules. These recombinant variant proteins can have altered biological activity.

Each individual protein that comprises the fusion protein can itself have altered biological activity compared to the activity of the wildtype protein. Altered biological activity is defined herein as activity different from that of the wildtype or recombinant protein. For

example, the activity of EPO is to regulate the growth and differentiation of red blood cell progenitors. Recombinant EPO variant proteins can have increased activity relative to wildtype EPO to regulate growth and differentiation of red blood cell progenitor cells. Alternatively, the EPO variant proteins containing mutations in the noncoding regions of the gene (e.g., 3' and 5' untranslated regions) can have decreased biological activity relative to the wildtype EPO.

Mutations in noncoding regions of the gene (e.g., 5' untranslated regions or UTR) can lead to differences in RNA translation as described, e.g., in Schultz, D.E., *et al.*, *J. Virol.* 70:1041-1049, 1996; Kozak, M., *J. Mol. Biol.* 235:95-110, 1994; and Kozak, M., *J. Biol. Chem.* 266:19867-19870, 1991. For example, as described in detail in Example 4, computer modeling can be used to predict differences in RNA secondary structure (e.g., free energy of loops and base pairs) following nucleotide alterations in 3' and 5' UTR of the EPO gene. Although secondary structure changes in EPO RNA, following mutations in the 5' or 3' UTR, are used as the specific example, it is understood that the instant invention described herein can be used to produce any suitable polypeptide variant protein. As used herein, the term mutation refers to any alteration in the nucleic acid sequence encoding a polypeptide (e.g., a point mutation; the addition, deletion and/or substitution of one or more nucleotides).

Secondary structure has been shown to be a critical component in determining the rates of translation efficiency of several proteins (Bettany, A.J., *et al.*, *J. Biol. Chem.* 267:16531-16537, 1992; Kozak, M., *J. Mol. Biol.* 235:95-110, 1994). By implication, altered rates of translation can affect posttranslational modifications, for example, glycosylation patterns, and, thus, proper folding of the resulting protein leading to changes in the chemistry, structure and function of the protein. The recombinant variant proteins described herein are unique in that they are composed of fusion proteins produced by mutations in 5' and 3' untranslated (noncoding) regions of the gene.

The present invention also provides recombinant expression vectors which include synthetic or cDNA-derived DNA fragments encoding fusion proteins comprising DNA encoding two or more linked proteins operably linked to suitable transcriptional or translational regulatory elements derived from mammalian, microbial, viral or insect genes. Such regulatory elements include a transcriptional promoter, an optional operator sequence to control transcription, a sequence encoding suitable mRNA ribosomal binding sites, and sequences which control the termination of transcription and translation, as described in detail below. The ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants may additionally be incorporated. Operably linked indicates that components

are linked in such a manner that expression of the DNA encoding a fusion protein is controlled by the regulatory elements. Generally, operably linked means contiguous.

Transformed host cells are cells into which fusion protein vectors have been introduced by infectious or non-infectious methods. Transformed host cells ordinarily express the desired fusion protein, but host cells transformed for purposes of cloning or amplifying DNA do not need to express the protein. In eukaryotic cells, expressed fusion protein will generally be secreted into the culture supernatant. In prokaryotic cells, the fusion proteins may be expressed within the periplasmic space or as insoluble inclusion bodies. Suitable host cells for expression of fusion protein include prokaryotes, yeast or higher eukaryotic cells under the control of appropriate promoters. Prokaryotes include gram negative or gram positive organisms, for example *E. coli*. Higher eukaryotic cells include established cell lines of mammalian origin as described below. Cell-free translation systems could also be employed to produce fusion protein using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described by Pouwels *et al.* ("*Cloning Vectors: A Laboratory Manual*", Elsevier, NY, 1985); Sambrook *et al.*, "*Molecular Cloning: A Laboratory Manual*", Second Edition (1989); and Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, Inc. (1997), which are

incorporated herein in their entirety by reference. Such techniques would be familar to one of skill in the art.

Prokaryotic expression vectors generally comprise one or more phenotypic selection markers, for example a gene encoding proteins conferring antibiotic resistance or supplying an autotrophic requirement, and an origin of replication recognized by the host to ensure amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium*, and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

Useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well-known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotech, Madison, WI). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. *E. coli* is typically transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species (Bolivar *et al.*, *Gene* 2:95, 1977). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides simple means for identifying transformed cells.

Promoters commonly used in recombinant microbial expression vectors include the blactamase

(penicillinase)and lactose promoter system (Chang *et al.*, *Nature* 275:615, 1978; and Goeddel *et al.*, *Nature* 281:544, 1979), the tryptophan (trp) promoter system (Goeddel *et al.*, *Nucleic Acids Res.* 8:4057, 1980) and tac promoter (Sambrook, *et al.*, "*Molecular Cloning: A Laboratory Manual*", 1989).

Recombinant fusion proteins may also be expressed in yeast hosts, preferably from the Saccharomyces species, such as *S. cerevisiae*. Yeast of other genera such as Pichia or Kluyveromyces may also be employed. Yeast vectors will generally contain an origin of replication from a yeast plasmid or an autonomously replicating sequence (ARS), promoter, DNA encoding the fusion protein, sequences for polyadenylation and transcription termination and a selection gene. Preferably, yeast vectors will include an origin of replication and selectable marker permitting transformation of both yeast and *E. coli*, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* trp1 gene, which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, and a promoter derived from a highly expressed yeast gene to induce transcription of a structural sequence downstream. The presence of the trp1 lesion in the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoter sequences in yeast vectors include the promoters for metallothionein, 3-phosphoglycerate

kinase (Hitzeman, *et al.*, *J. Biol. Chem.* 255:2073, 1980) or other glycolytic enzymes (Hess, *et al.*, *J. Adv. Enzyme Reg.* 7:149, 1968; and Holland, *et al.*, *Biochem.* 17:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose-phosphate isomerase, phosphoglucose isomerase, and glucokinase. Suitable vectors and promoters for use in yeast expression are further described in R. Hitzeman, *et al.*, EPA 73,657.

Preferred yeast vectors can be assembled using DNA sequences from pBR322 for selection and replication in *E. coli* (Amp gene and origin of replication) and yeast DNA sequences including a glucose-repressible ADH2 promoter and α-factor leader, which directs secretion of heterologous proteins, can be inserted between the promoter and the structural gene to be expressed (Kurjan, *et al.*, *Cell* 30:933, 1982; and Bitter, *et al.*, *Proc. Natl. Acad. Sci. USA* 81:5330, 1984). The leader sequence may be modified to contain, near its 3' end, one or more useful restriction sites to facilitate fusion of the leader sequence to foreign genes.

Suitable yeast transformation protocols are known to those of skill in the art; an exemplary technique is described by Hinnen, *et al.*, *Proc. Natl. Acad. Sci. USA* 75:1929, 1978, selecting for Trp^+ transformants in a selective medium consisting of 0.67% yeast nitrogen base,

0.5% casamino acids, 2% glucose, 10 µg/ml adenine and 20 µg/ml uracil.

Host strains transformed by vectors comprising the ADH2 promoter may be grown for expression in a rich medium consisting of 1% yeast extract 2% peptone, and 1% glucose supplemented with 80 µg/ml adenine and 80 µg/ml uracil. Derepression of the ADH2 promoter occurs upon exhaustion of medium glucose. Crude yeast supernatants are harvested by filtration and held at 4°C. prior to further purification. Various mammalian or insect cell culture systems can be employed to express recombinant protein. Baculovirus systems for production of heterologous proteins in insect cells are reviewed by Luckow and Summers, *Bio/Technology* 6:47, 1988.

Examples of suitable mammalian host cell lines include the COS-7 lines of monkey kidney cells, described by Gluzman (*Cell* 23:175, 1981), and other cell lines capable of expressing an appropriate vector including, for example, L cells, C127, 3T3, Chinese Hamster Ovary (CHO), HeLa and BHK cell lines. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter and enhancer linked to the gene to be expressed, and other 5' or 3' flanking nontranscribed sequences, and 5' to 3' nontranslated sequences, such as necessary ribosome binding sites, a poly-adenylation site, splice donor and acceptor sites, and transcriptional termination sequences.

The variant nucleic acid molecules encoding, for example, recombinant polypeptide variant proteins created by modifying the 3' and/or 5' UTR of the polypeptide gene, would also preferably contain regulatory sequences. Regulatory sequences include all cis-acting elements that control transcription and regulation such as, promoter sequences, enhancers, ribosomal binding sites, and transcription binding sites. Selection of the promoter will generally depend upon the desired route for expressing the protein. For example, where the protein is to be expressed in a recombinant eukaryotic or prokaryotic cell, the selected promoter is recognized by the host cell. A suitable promoter which can be used can include the native promoter for the binding moiety which appears first in the construct.

The elements which comprise the nucleic acid molecule can be isolated from nature, modified from native sequences or manufactured *de novo*, as described, for example, in the above-referenced texts. The elements can then be isolated and fused together by methods known in the art, such as exploiting and manufacturing compatible cloning or restriction sites.

The nucleic acid molecules can be inserted into a construct which can, optionally, replicate and/or integrate into a recombinant host cell, by known methods which may vary depending upon the form of the recombinant polypeptide which is expressed. The host cell can be a eukaryotic or prokaryotic cell and includes, for example, pichia

expression systems, yeast (such as, Saccharomyces), bacteria (such as, Escherichia or Bacillus), animal cells or tissue, including insect (such as, *Spodoptera frugiperda 9* or mammalian cells (such as, somatic or embryonic human cells, Chinese hamster ovary cells, HeLa cells, human 293 cells, monkey kidney COS-7 cells, baby hamster kidney BHK cells, C127 cells, etc.). The selection of the host cell governs the posttranslational modifications that may occur. For instance, glycoproteins could be expressed in mammalian, insect, or yeast cells whereas nonglycosylated proteins could be expressed in bacteria. In addition, the selection of the appropriate host cell may differ when expressing recombinant polypeptide variants manufactured by mutations in the noncoding regions of the gene. (Schultz, *et al.*, *J. Virol.* 70:1041-1049, 1996).

The nucleic acid molecule can be incorporated or inserted into the host cell by known methods. Examples of suitable methods of transfecting or transforming cells include calcium phosphate precipitation, electroporation, microinjection, infection, lipofection and direct uptake. Methods for preparing such recombinant host cells are described in more detail in several technical books, for example, Sambrook, *et al.*, "Molecular Cloning: A Laboratory Manual," (1989) and Ausubel, *et al.* "Current Protocols in Molecular Biology," (1995).

The host cells are then maintained under suitable conditions for expressing and recovering the recombinant polypeptide. Generally, the cells are maintained in a

suitable buffer and/or growth medium or nutrient source for growth of the cells and expression of the gene product(s). The growth media are generally known in the art and include sources of carbon, nitrogen and sulfur. Examples include Dulbeccos modified Eagles media (DMEM), RPMI-1640, M199 and Grace's insect media. The selection of a buffer is not critical to the invention. The pH which can be selected is generally one tolerated by or optimal for growth for the host cell.

The cell is maintained under a suitable temperature and atmosphere. For example, an aerobic host cell is maintained under aerobic atmospheric conditions or other suitable conditions for growth. The temperature should also be selected so that the host cell tolerates the process and can be, for example, between about 27°C and 40°C.

The transcriptional and translational control sequences in expression vectors to be used in transforming vertebrate cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from Polyoma, Adenovirus 2, Simian Virus 40 (SV40), and human cytomegalovirus. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide the other genetic elements required for expression of a heterologous DNA sequence. The early and late promoters are particularly useful because both are obtained easily from the virus as a fragment which also

contains the SV40 viral origin or replication (Fiers, *et al.*, *Nature* 273:113, 1978). Smaller or larger SV40 fragments may also be used, provided the approximately 250 bp sequence extending from the Hind III site toward the BgII site located in the viral origin or replication is included. Exemplary vectors can be constructed as disclosed by Okayama and Berg (*Mol. Cell. Biol.* 3:280, 1983).

Preferred eukaryotic vectors for expression of mammalian DNA include pIXY321 and pIXY344, both of which are yeast expression vectors derived from pBC102.K22(ATCC 67,255) and contain DNA sequences from pBR322 for selection and replication in *E. coli* (Apr gene and origin of replication) and yeast.

Purified mammalian fusion proteins or analogs are prepared by culturing suitable host/vector systems to express the recombinant translation products of the DNAs of the present invention, which are then purified from culture media or cell extracts. For example, supernatants from systems which secrete recombinant protein into culture media can be first concentrated using a commercially available protein concentration filter, for example, an Amicon or Millipore Pellicon ultrafiltration unit. Following the concentration step, the concentrate can be applied to a suitable purification matrix.

The recombinant molecules produced by the processes described herein, including those manufactured by modifications in the 3' and 5' UTR, can be isolated and

purified by known means. Examples of suitable purification and isolation processes are generally known in the art and include, but are not limited to, ammonium sulfate precipitation, dialysis, electrophoresis, ultrafiltration, microfiltration, gel filtration, ion exchange or immunoaffinity chromatography. In addition, one or more reverse phase high performance liquid chromatography (RP-HPLC) media, e.g., silica gel having pendant methyl or other aliphatic groups, can be employed to further purify a fusion protein composition. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a homogenous recombinant protein.

Recombinant protein produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more concentration, salting-out, aqueous ion exchange or size exclusion chromatography steps. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of recombinant fusion proteins can be disrupted by any convenient method, including freezethaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Fermentation of yeast which express fusion proteins as a secreted protein greatly simplifies purification. Secreted recombinant protein resulting from a large scale fermentation can be purified by methods analogous to those disclosed by Urdal, *et al.*, (*J. Chromatog.* 296:171, 1984).

Fusion protein synthesized in recombinant culture is characterized by the presence of non-human cell components, including proteins, in amount and of a character which depend upon the purification steps taken to recover the fusion protein from the culture. These components ordinarily will be of yeast, prokaryotic or non-human higher eukaryotic origin and preferably are present in innocuous contaminant quantities, on the order of less than about 5 percent by scanning densitometry or chromatography. Further, recombinant cell culture enables the production of the fusion protein free of proteins which may be normally associated with EPO as they are found in nature in their respective species of origin, e.g., in cells, cell exudates or body fluids.

The present invention further relates to pharmaceutical compositions comprising a fusion protein and a physiologically-compatible carrier. Such carriers are well-known, routine and described in U.S. Patent No. 5,580,853, the teachings of which are herein incorporated by reference in their entirety. Pharmaceutical compositions suitable for administration comprise the fusion protein in an effective amount and a physiologically acceptable carrier.

An effective amount, as used herein, is defined as that quantity which alleviates, to any degree, or eliminates the condition for which the mammal is being treated.

The carriers will be non-toxic to recipients at the dosages and concentrations employed. The formulation used will vary according to the route of administration selected (e.g., solution, emulsion, capsule). For solutions or emulsions, suitable carriers include, for example, aqueous or alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles can include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles can include various additives, preservatives, or fluid, nutrient or electrolyte replenishers. See, generally, "*Remington's Pharmaceutical Science*", 16th Edition, Mack, Ed. (1980). For inhalation, the compound can be solubilized and loaded into a suitable dispenser for administration (e.g., an atomizer, nebulizer or pressurized aerosol dispenser). Fusion proteins can be administered individually, together or in combination with other drugs or agents (e.g., other chemotherapeutic agents, immune system enhancers).

Fusion protein compositions can be used to enhance proliferation, differentiation and functional activation of hematopoietic progenitor cells, such as bone marrow cells. Fusion protein compositions can also be used in the treatment of cancers or cellular growth deficiencies. Specifically, compositions containing the fusion protein may be used to increase peripheral blood leukocyte numbers and increase circulating granulocyte counts in myelosuppressed patients. To achieve this result, a

therapeutically effective quantity of a fusion protein composition is administered to a mammal, preferably a human, in association with a pharmaceutical carrier or diluent.

The recombinant polypeptide variant proteins of the invention can be used as therapeutic for delivery to individuals having diseases or conditions that are associated with deficiencies or abnormalties of the proteins described herein. The retention and/or deletion of nucleotides in untranslated regions of the polypeptide gene can produce heterologous therapeutic proteins. Heterologous proteins are herein defined as proteins which does not exist in nature and exhibit a range of therapeutic effects.

Recombinant polypeptides with therapeutic value are known in the art. Examples include the recombinant EPO described in Lin (U.S. Patent No. 4,703,008); Sytkowski *et al.*, (U.S. Patent No. 5,614,184); Sytkowski (U.S. Patent No. 5,580,853); and Powell (U.S. Patent No. 5,688,679); the contents of which are incorporated herein by reference. Therapeutic advantages of recombinant EPO include, for example, an increase in hematocrit and specific activity of the recombinant protein. A recombinant EPO protein molecule with increased biological activity can be generated by substitution of glycine at position 101 with another amino acid residues such as alanine (Sytkowski *et al.*, U.S. Patent No. 5,614,184 (1997)). These references can be used to guide one of skill in the art in the

manufacturing of nucleic acid constructs encoding protein molecules which comprise the fusion proteins of the invention.

For example, the recombinant EPO variant proteins described herein can be employed in any method where EPO would be effective, and in particular in methods where other man-made EPO proteins have not produced any clinically beneficial effect (e.g., increasing red blood cells in an anemic patient). The mode of EPO administration to patients is preferably at the location of the target cells. As such, the administration can be by injection. Other modes of administration (parenteral, mucosal, systemic, implant, intraperitoneal, etc.) are generally known in the art and, for EPO, can be determined, for example, as described in U.S. Patent No. 5,614,184. The recombinant EPO proteins can, preferably, be administered in a pharmaceutically acceptable carrier, such as saline, sterile water, Ringer's solution, and isotonic sodium chloride solution.

The activity of polypeptide variant proteins can be tested, for example, in pharmacological differences. Accordingly, the activity of the recombinant EPOs could be evaluated therapeutically. For example, pharmacological differences in the secreted and purified EPO manufactured by the disclosed method compared to other man-made or naturally occurring EPOs can include:

1. An increase or decrease in the potency when administered to patients in human clinical

trials. The difference can be in the required initial dose as well as maintenance doses. A relative potency factor can be evaluated for the recombinant EPO variant proteins.

2. A reduction or increase in potential side effects in patients may reflect altered activities of the EPO variant proteins. For example, differences can be manifested as an increase or decrease in blood pressure which can be of extraordinary significance in designing treatment regimens for certain high risk patients like dialysis patients who are, in any case, severely ill.

3. A difference in the time lag between the effect of increasing red blood cells in the patient's serum after administration of the EPO variants. This time-lag has the consequence that the desired therapeutic effect is either accelerated or delayed significantly compared to other forms of recombinant EPO. A decrease in the time lag would be a desirable therapeutic effect by resulting in a faster benefit to the patient.

4. The ability of a patient to tolerate one form of EPO and not another. If a patient can not tolerate one form of an EPO variant over another, this noncompatibility can indicate therapeutic differences which in turn can reflect structural, biochemical and biological modifications in the various forms of recombinant EPO.

5. An increase in the circulating half-life of EPO in patients which can result in less frequent injections or smaller doses of EPO having to be administered. A prolonged half-life would not only be therapeutically beneficial, but also diminish health care costs in the treatment of chronically ill patients.

Thus, differences in the pharmaceutical characteristics of recombinant polypeptide variant proteins can result in variations in therapeutic effects (e.g., for EPO variants, the production of reticulocytes and red blood cells and an increase in hemoglobin synthesis and iron uptake). For example, a difference in the inherent potency which would result in lower bioloads inflicted on the patient's body by administering an EPO protein which leads to an absence or drastic lowering of side effects (which may endanger the patient's life or make it impossible to administer one form of EPO) is particularly important in high risk patients (e.g., patients with kidney disorders) who are at high risk for hypertension, myocardial infarct or stroke.

Thus, retention, deletion, point mutation or substitution in the 5' and/or 3' UTR sequences of a recombinant EPO gene fragment can ultimately influence the final structure and chemistry of a protein expressed by a host cell transfected with that gene fragment. As a consequence the resulting expressed protein can exhibit

varying biological parameters which can be assessed using bioassays and in therapeutics.

The present invention will now be further illustrated by the following exemplification, which is not meant to be limiting in any way.

EXAMPLE 1

CONSTRUCTION OF THE EPO-EPO DIMER

An EPO-EPO fusion protein (Figures 16A-16C; SEQ ID NOS: 16 and 17) was constructed by linking two strands of EPO cDNA with a DNA strand encoding the following polypeptide: AGGGGSGGGGSGGGGST (SEQ ID NO: 18) (Figure 1). The nucleotide sequence of wild type erythropoietin is described in from Jacobs, K., *et al.*, *Nature* 323:806, 1985, which is herein incorporated by reference in its entirety. The linking DNA strand was sequentially lengthened to the proposed length by using psv2-EPO (Figure 2) as template and 3' primers with appropriately extended 3' ends (Figure 2). The initial preceding EPO DNA strand (Figure 4) contains 10 nucleotides in the 5' untranslated region, a leader sequence (Jacobs, K., *et al.*, *Nature* 323:806, 1985), an EPO cDNA coding sequence, and no STOP codon. Additional nucleotides attached to the 3' end were: GCCGGCGGTGGTGGATCTGG (SEQ ID NO: 19). The EPO DNA strand after the linker (EPO B DNA; Figure 5) contains no leader sequence but has a STOP codon and 17 nucleotides in the 3' untranslated region. Half a NaeI restriction site was designed into the 3' end of EPO A

DNA and half a ScaI restriction site into the 5' end of EPO B DNA.

EPO A (Figure 2) and EPO B (Figure 3) DNA were produced by the using the Polymerase Chain Reaction (PCR) and a human EPO cDNA plasmid, psv2-EPO (Chern, Y.J., *et al.*, *Eur J Biochem* 202:225, 1991) as template.

Primers used to produce EPO A are as follows: 5'-AGGCGCGGAGATGGGGGTGCAC (SEQ ID NO: 20) (EpA 5'), 3'-CCAGATCCACCACCGCCGGCTCTGTCCCCTGTCCTGCAGG (SEQ ID NO: 21) (EpA3-3), 3'-CGCCACCGGATCCACCGCCACCAGATCCACCACCGCCGGC (SEQ ID NO: 22) (EpA3-4), and 3'-TGGTGGGGCAGTACTGCCGCCGCCACCGGATCCACCGCC (SEQ ID NO: 23) (EpA3-5).

Primers used to produce EPO B are as follows: 5'-GCGGCAGTACTGCCCCACCACGCCTCATCTGTGACAGC (SEQ ID NO: 24) (EpB 5-1) and 3'-CAGGTGGACACACCTGGTCATC (SEQ ID NO: 25) (EpB 3').

PCR reactions (50 µl)contained the following components: 0.5 µM of 5' primer or 3' primer; 10 ng psv2-EPO; 200 µM of dATP, dCTP, dGTP, or dTTP; 20 mM Tris-HCl (pH 8.0); 2 mM $MgCl_2$; 10 mM KCl; 6 mM $(NH_4)_2SO_4$; 0.1% Triton X-100; 10 µg/ml nuclease-free BSA; and 2.5 U Pfu DNA Polymerase (Stratagene). The reactions were overlaid with mineral oil (50 µl; Molecular Biology Grade, Sigma) and subjected to 25 cycles of 94°C for 1 min (denaturation), of 52°C for 1 min (annealing) and of 72°C for 1 min (extension) in a Perkin Elmer DNA Thermal Cycler 480.

Next, the DNA sequences of the PCR products were determined. First, the PCR products were purified from a 1% agarose gel using the QIAQUICK™ Gel Extraction Kit. They were then ligated to pCR-blunt, in which the reactions contained an insert to vector molar ratio of 10 to 1. The ligation reactions (10 µl) contained the gel-purified PCR products, 25 ng of PCR-blunt, 1X ligation buffer and 4 U of T4 DNA ligase (ZERO BLUNT™ PCR Cloning Kit, Invitrogen). Incubations were carried out for 1 hour at 16°C.

Cells used for expression were TOP 10™ Competent Cells (Invitrogen) and were transformed according to procedure established by Invitrogen: 2 µl of β-mercaptoethanol was added to the cells on ice, mixed by gentle swirling with a pipette tip, followed by 2 µl of the ligation described in the preceding paragraph. This mixture was then incubated on ice for 30 min, followed by exactly 45 seconds at 42°C. The vial was then placed on ice for 2 min. Pre-warmed (37°C) SOC medium (250 µl) containing 2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose was added and the cells shaken for 1 hr at 37°C. Fifty µl of a 1:5 dilution of transformed cells were plated on LB (Miller's modification, Sigma) agar plates containing 50 µg/ml kanamycin. The plates were incubated at 37°C overnight. Colonies were plucked and 2.5 ml LB containing 50 µg/ml kanamycin were inoculated with these colonies. Plasmid DNA were prepared from the overnight cultures using Promega's WIZARD PLUS MINIPREPS™ DNA Purification System.

Clones were analyzed by restriction digest fragment analysis.

The pCRBlunt-EPO A and pCRBlunt-EPO B DNA clones were digested with BglI, which gave unique-sized fragments for a correctly inserted DNA and an insert oriented in the reverse direction (Figures 6 and 7). Clones with inserts in the reverse direction were chosen and larger amounts (from 100 ml of LB/50 µg/ml kanamycin) of DNA plasmids prepared using Promega's WIZARD PLUS MAXIPREPS™ DNA purification system. Clones with inserts in the "forward" direction would also have produced the proposed EPO-EPO DNA.

EPO A DNA was linked to EPO B DNA using the procedure described as follows. pCRBlunt-EPO A(-) was digested with Sca I and Xho I and the 677 bp fragment gel purified (Figure 8). pCRBlunt-EPO B(-) was digested with BamHI and ScaI and the 557 bp fragment gel purified (Figure 9). The EPO A 677 bp fragment was then ligated to the EPO B 557 bp fragment in a 1:1 molar ratio of EPO A 677 bp fragment to EPO B 557 bp fragment. Ligations were carried out overnight at 16°C. The ligated EPO A/EPO B DNA fragments were purified using QIAQUICK™ Gel Extraction Kit then ligated to pcDNA2.1(-) which had previously been digested with XhoI and BamHI and gel purified (Figure 10). The ligation reaction contained a 5:1 molar ratio of DNA insert to pcDNA3.1(-). The incubation was carried out overnight at 16°C. Clones were picked from ampicillin-resistant colonies by restriction digest analysis (Figure 11),

produced in microgram quantities, and used to transfect COS1 cells.

EXAMPLE 2

TRANSIENT EXPRESSION OF EPO DIMER IN COS1 CELLS

COS1 cells were grown to 70% confluency in Dulbecco's Modified Eagle Medium, high glucose (4.5 g/L; Gibco), 10% fetal bovine serum (Hyclone) in the presence of 100 U penicillin, 100 µg streptomycin, 250 ng Fungizone per ml of tissue culture medium (antibiotic-antimycotic cocktail from Gibco) at 37°C and 10% CO_2. The cells were harvested by trypsinizing using 0,05% trypsin, 0.53 mM EDTA (Gibco) and washing twice with phosphate buffered saline (PBS)/6 mM glucose solution. Cells were suspended in the above PBS/glucose buffer to a concentration of 2 x 10^6 cells/ml. 0.5 ml of cells were placed in electroporation cuvettes (0.4 cm gap, Bio-Rad) and 10 µg of pcDNA/EPO-EPO added. The cells were electroporated under the following conditions: voltage = 0.3 kV, field strength = 0.75 kV/cm, capacitor = 250 µF, and resistor = none (Pulse controller set at Ω). Cells were plated in 30 ml of pre-warmed DMEM, high glucose, 10% FBS and incubated for 72 h at 37°C and 10% CO_2. The controls used were 10 µg of pcDNA-EPO and 10 µg of pcDNA 3.1(-). Transfected and non-transfected cells were cultured for 3 days before analysis.

Total RNA from transiently transfected COS1 cells was prepared using TRIZOL Reagent (GibcoBRL) according to the manufacturer's protocol. Total cellular RNA was separated

produced in microgram quantities, and used to transfect COS1 cells.

EXAMPLE 2

TRANSIENT EXPRESSION OF EPO DIMER IN COS1 CELLS

COS1 cells were grown to 70% confluency in Dulbecco's Modified Eagle Medium, high glucose (4.5 g/L; Gibco), 10% fetal bovine serum (Hyclone) in the presence of 100 U penicillin, 100 µg streptomycin, 250 ng Fungizone per ml of tissue culture medium (antibiotic-antimycotic cocktail from Gibco) at 37°C and 10% CO_2. The cells were harvested by trypsinizing using 0,05% trypsin, 0.53 mM EDTA (Gibco) and washing twice with phosphate buffered saline (PBS)/6 mM glucose solution. Cells were suspended in the above PBS/glucose buffer to a concentration of 2 x 10^6 cells/ml. 0.5 ml of cells were placed in electroporation cuvettes (0.4 cm gap, Bio-Rad) and 10 µg of pcDNA/EPO-EPO added. The cells were electroporated under the following conditions: voltage = 0.3 kV, field strength = 0.75 kV/cm, capacitor = 250 µF, and resistor = none (Pulse controller set at Ω). Cells were plated in 30 ml of pre-warmed DMEM, high glucose, 10% FBS and incubated for 72 h at 37°C and 10% CO_2. The controls used were 10 µg of pcDNA-EPO and 10 µg of pcDNA 3.1(-). Transfected and non-transfected cells were cultured for 3 days before analysis.

Total RNA from transiently transfected COS1 cells was prepared using TRIZOL Reagent (GibcoBRL) according to the manufacturer's protocol. Total cellular RNA was separated

on 1.2% agarose gels containing 5.5% formaldehyde and transferred to a GeneScreen Plus filter. The filter was probed with a digoxigenin-dUTP Boehringer Mannheim according to the procedure described by the manufacturer. The EPO-L-EPO fusion protein was encoded by a transcript of approximately 2.8 kb in length. The relative amount of EPO-L-EPO transcript, as assessed by Northern blot analysis, was approximately equivalent to that observed for EPO monomer. Since the secretion of EPO-L-EPO protein was at least eight fold higher than EPO (See Examples 3 and 4), it is likely that the EPO-L-EPO protein dimer is considerably more stable during synthesis than the EPO monomer.

The conditioned media from transfected and nontransfected cells were collected, centrifuged at 13,800 x g for 10 min at 4°C. Aliquots (1 ml) of each conditioned media were dialyzed against Minimum Essential Medium α overnight with 3 changes of medium. The dialyzed samples were assayed for the determination of EPO protein concentrations and used for evaluating *in vitro* and *in vivo* EPO activity.

EXAMPLE 3

EPO PROTEIN DETERMINATION IN COS1 CELL CULTURE MEDIUM

EPO monomer and EPO-L-EPO fusion protein secreted by transfected COS1 cells were detected by Western blotting and Enzyme-Linked Immunoabsorbant Assay (ELISA) in dialyzed conditioned culture media from transfected COS1 cells.

Proteins present in dialyzed samples were separated by SDS-PAGE and electrophoretically transferred to 0.45 μm nitrocellulose membranes in a buffer containing 25 mM Tris-HCl, 192 mM glycine, 10% methanol. Membranes were then briefly rinsed twice with distilled water and incubated overnight at 4°C in 20 mM Tris-HCl, 0.5 M NaCl, 0.5% Tween-20 (TBST), 10% nonfat dry milk, pH 7.5. The membranes were rinsed twice with TBST, washed once with TBST for 15 min and twice for 5 min each. The membranes were then incubated for 1 hr at 23°C with anti-erythropoietin monoclonal antibody AE-7A5 (Genzyme, Corp, Cambridge, MA) at a concentration of 0.7 μg/ml in TBST containing 5% nonfat dry milk. Rinsing and washing were performed as described above, followed by incubation for 1 hr at 23°C with a horse-radish peroxidase conjugated goat anti-mouse IgG (Cappel) diluted 1:1000 in TBST containing 5% nonfat dry milk. Rinsing and washing were again performed as described above except that two additional TBST washes (5 min each) were performed. Antigen (EPO)-antibody complexes were visualized by chemiluminescence detection using an Amersham ECL kit. The EPO-L-EPO dimer was secreted from COS1 cells as a single protein band with a molecular weight of 76 kDa. The molecular weight of the secreted EPO dimer is approximately equivalent to the molecular weight of two fully processed and glycosylated EPO monomers (36 kDa) and the polypeptide linker (1.8 kDa). Therefore, the dimer consists of two complete, mature and fully processed EPO protein molecules.

An EPO ELISA assay (Genzyme Corp., Cambridge, MA) was used to determine the amount of EPO monomer or dimer present in conditioned media of transfected COS1 cells. The standard curve was calculated using recombinant monomeric EPO (rhuEPO) according to routine methods described previously. See, for example, Sytkowski *et al.*, U.S. Patent No. 5,614,184 (1997) and Sytkowski *et al.*, *Proc. Natl. Acad. Sci. USA 95*:1184-1188 (1998).

The EPO-L-EPO fusion protein dimer was secreted from COS1 cells at a concentration of 0.15 µg/ml which is approximately eight-fold higher than the amount EPO monomeric protein secreted by COS1 cells (0.018 µg/ml) (See Table 1, Example 4). As discussed above, it is possible that the protein dimer is considerably more stable than the monomer protein molecule during protein synthesis. Such a property can be advantageous for *in vitro* production and *in vivo* therapeutic purposes.

EXAMPLE 4

IN VITRO BIOACTIVITY OF EPO DIMER

The *in vitro* activity of EPO-L-EPO dimer protein was determined according to the method of Krystal (*Exp. Hematol. 11*:649-660 (1983)) as described herein and as described in U.S. Patent No. 5,580,853, which is herein incorporated by reference in its entirety. As previously discussed the Krystal bioassay measures the effect of EPO monomer or EPO fusion proteins on erythroid cell proliferation using intact mouse spleenic erythroid cells.

Recombinant EPO was used to generate the standard curve against the World Health Organization Second International Reference Preparation. Each sample was diluted in assay medium containing 78% α-MEM, 20% heat inactivated fetal bovine serum, 1% β-mercaptoethanol and 1% penicillin/streptomycin/fungizone. The assay is routine and familiar to one of skill in the art.

The conditioned medium of COS1 cells transfected with pcDNA/EPO-EPO contained *in vitro* bioactivity of 45-72 U/ml, and the medium from COS1 cells transfected with pcDNA-EPO contained 5 U/ml. Conditioned media from cells transfected with pcDNA and untransfected cells showed no EPO activity.

The specific activity (U/μg) of EPO dimer and monomer was calculated by dividing the *in vitro* bioassay value (U/ml) by the concentration of protein as determined by ELISA (μg/ml). As shown in Table 1 the specific activity of the EPO^{wt}-L-EPO^{wt} dimer was 1007 U/ml compared to EPO monomer with a specific activity of 350 U/ml. Thus, dimeric proteins comprising wildtype EPO are secreted from COS1 cells at enhanced rates and have increased biological and specific activity compared to monomeric wildtype EPO, in excess of twice the amount that would be predicted for a fusion protein composed of two units.

TABLE 1 BIOLOGICAL ACTIVITY AND PROTEIN CONCENTRATION IN TRANSFECTED COS1 CELL MEDIUM

Construct	Protein secreted		Specific Activity
	Bioassay U/ml	ELISA µg/ml	U/µg
Epo^{wt}	6.3	0.018	350
Epo^{wt}/Epo^{wt}	151	0.150	1007
Epo^{R103A}/Epo^{wt}	135	0.281	480
Epo^{wt}/Epo^{R103A}	123	0.238	516
Epo^{R103A}/Epo^{R103A}	ND	ND	---

ND = nondetectable

EXAMPLE 5

IN VIVO ACTIVITY OF EPO-EPO DIMERS

Conditioned medium from COS1 cells transfected with pcDNA/EPO-EPO was used to inject mice (B6C3F1 strain, female, 18 g, Jackson Labs). The haematocrits of these mice were measured prior to administering EPO-EPO (Pre-treatment values). Mice were injected subcutaneously (EPO-EPO 300 IU per kg) on day 1, day 3, and day 5 and post-treatment haematocrits determined 7 days later in day 8 of the experimental treatment regimen. Mouse #1 showed an increase haematocrit of 4.5% and mouse #2 an increase of 1.5%.

In another more extensive series of experiments, mice received a single injection on day 1 of the treatment regimen of conditioned media from COS1 cells transfected

with either pcDNA/EPO-EPO (n=4 mice) or pcDNA/EPO (n=4 mice). As a control, another group of mice (n=4 mice) received a single dose of conditioned media from COS1 cells. Post-treatment hemotocrits were determined 7 days later or on day 8 of the treatment regimen and compared to pre-treatment (day 0) values.

As shown in Figures 17A, 17B, and 17C a substantial increase in hematocrits, compared to pre-treatment values, was observed only in mice injected with conditioned media containing the EPO-EPO dimer. No increase in hematocrit was observed in mice treated with EPO monomer or with control culture media. These *in vivo* data suggest that the EPO-EPO dimer has a prolonged plasma half-life.

EXAMPLE 6

OLIGONUCLEOTIDE-DIRECTED MUTAGENESIS OF FUSION PROTEINS

Mutations in nucleic acid constructs encoding the fusion proteins of the invention, for example EPO, can be prepared using oligonucleotide-directed mutagenesis according to the Altered Sites™ *In Vitro* Mutagenesis System kit (Promega Corporation of Madison, WI). The Altered Sites™ System is based on experimental protocols routine to one of ordinary skill in the art of molecular biology. The kit consists of a unique mutagenesis vector and a simple, straightforward procedure for selection of oligonucleotide-directed mutants. The system is based on the use of a second mutagenic oligonucleotide to confer antibiotic resistance to the mutant DNA strand. The system employs a

phagemid vector, pSELECT-1, which contains two genes for antibiotic resistance. One of these genes, for tetracycline resistance, is always functional. The other, for ampicillin resistance, is inactivated. An oligonucleotide is provided which restores ampicillin resistance to the mutant strand during the mutagenesis reaction. This oligonucleotide is annealed to the single-stranded DNA (ssDNA) template at the same time as the mutagenic oligonucleotide and subsequent synthesis and ligation of the mutant strand links the two. The DNA is transformed into a repair minus strain *E. coli*, or other suitable host, and the cells are grown in the presence of ampicillin, yielding large numbers of colonies. A second round of transformation in JM109, or a similar host, ensures proper segregation of mutant and wild type plasmids and results in a high proportion of mutants.

The pSELECT-1 plasmid is a phagemid, defined as a chimeric plasmid containing the origin of a single-stranded DNA bacteriophage. This phagemid produces ssDNA upon infection of the host cells with the helper phage R408 or M13KO7. The vector contains a multiple cloning site flanked by the SP6 and T7 RNA polymerase promoters and is inserted into the lacZ α-peptide. Cloning of a DNA insert into the multiple cloning site results in inactivation of the α-peptide. When plated on indicator plates, colonies containing recombinant plasmids are white in a background of blue colonies. The SP6 and T7 promoters may be used to generate high specific activity RNA probes from either

strand of the insert DNA. These sites also serve as convenient priming sites for sequencing of the insert. The pSELECT-1 vector carries gene sequences for both ampicillin and tetracycline resistance. However, the plasmid is ampicillin sensitive because a frameshift is introduced into this resistance gene by removing the Pst I site. Therefore, propagation of the plasmid and recombinants is performed under tetracycline selection.

The pSELECT-Control vector provides a convenient white/blue positive control for mutagenesis reactions. This vector is derived from the pSELECT-1 vector by removing the Pst I site within the polylinker. The resultant frameshift in the lac α-peptide inactivated β-galactosidase and leds to a white colony phenotype on indicator plates. A lacZ repair oligonucleotide (supplied with the system) may be used to introduce a four base insertion which corrects the defect in the lacZ gene and restores colony color to blue. The fraction of blue colonies obtained is an indication of the mutagenesis efficiency. When the lacZ repair oligonucleotide is used in combination with the ampicillin repair oligonucleotide to correct this defect, 80-90% of the ampicillin resistant colonies are blue. When the lacZ repair oligonucleotide is used alone, a mutagenesis efficiency of only 2-5% is seen.

The mutagenic oligonucleotide must be complementary to the single-stranded target DNA. The ssDNA produced by the pSELECT-1 phagemid is complementary to the lacZ coding strand.

The stability of the complex between the oligonucleotide and the template is determined by the base composition of the oligonucleotide and the conditions under which it is annealed. In general, a 17-20 base oligonucleotide with the mismatch located in the center is sufficient for single base mutations. This provides 8-10 perfectly matched nucleotides on either side of the mismatch. For mutations involving two or more mismatches, oligonucleotides of 25 bases or longer are needed to allow for 12-15 perfectly matched nucleotides on either side of the mismatch.

Routinely, oligonucleotides are annealed by heating to 70°C for 5 minutes followed by slow cooling to room temperature.

DNA to be mutated is cloned into the pSELECT-1 vector using the multiple cloning sites. The vector DNA is then transformed into competent cells of JM109, or a similar host, and recombinant colonies are selected by plating on LB plates containing 15µg/ml tetracycline, 0.5mM IPTG, and 40µg/ml X-Gal. After incubation for 24 hours at 37°C, colonies containing recombinant plasmids appear white in a background of blue colonies.

To produce single-stranded template for the mutagenesis reaction, individual colonies containing pSELECT-Control or recombinant pSELECT-1 phagemids are grown and the cultures are infected with helper phage as described below. The single-stranded DNA produced is complementary to the lacZ coding strand and complementary

to the strand of the multiple cloning site. Two helper phages R408 and M13KO7 are used to provide the greatest latitude in optimizing ssDNA yields.

EXAMPLE 7

CONSTRUCTION AND EXPRESSION OF EPO-EPO MUTANT DIMERS

In order to determine whether one or both EPO domains of the EPO-EPO fusion protein are biologically active, mutant EPO-EPO dimers were constructed. A single point mutation (R103A) wherein the arginine at position 103 is replaced with an alanine residue in human recombinant monomeric EPO results in complete inactivation of the protein (Grodberg *et al., Eur. J. Biochem 218*:597-601 (1993); and Matthews *et al., Proc. Natl. Acad. Sci. USA 93*:9471-9476 (1996)). Therefore, the effect of the R103A mutation, in one or both of the monomeric subunits of the EPO-EPO dimer, on the biological activity of the EPO-EPO dimer was determined. Point mutations, including the R103A mutation, can be made using site-directed mutagenesis techniques as described above.

An EPO mutation (EPO^{R103A}) was created in the EPO nucleic acid construct encoding the N-terminal, or C-terminal, or both N-terminal and C-terminal domains of the dimer. EPO mutant dimers (EPO^{R103A}-EPO^{wt} ; EPO^{wt}-EPO^{R103A}; EPO^{R103A}-EPO^{R103A}) were generated using PCR and ligation reaction conditions essentially as described in Example 1. Likewise methods to transfect COS1 cells with mutant dimers and techniques to prepare conditioned media for evaluation

of biological activity were also as described above for EPO^{wt}-EPO^{wt} dimers.

EXAMPLE 8

BIOLOGICAL ACTIVITY OF EPO-EPO MUTANT DIMERS

As discussed above, mutation of the arginine at position 103 to an alanine residue in monomeric EPO leads to a complete loss of biological activity. Thus, it was expected that a R103A mutation in one domain of an EPO-EPO dimer (EPO^{R103A}-EPO^{wt} or EPO^{wt}-EPO^{R103A}) would result in a fusion protein which retained approximately half the biological activity of the EPO^{wt}-EPO^{wt} dimer. It was further expected that mutation of both domains (Epo^{R103A}/Epo^{R103A}) would result in complete inactivation of the fusion protein.

As shown in Table 1 (See Example 4) the EPO^{R103A}-EPO^{wt} and EPO^{wt}-EPO^{R103A} mutants were secreted from COS1 cells at concentrations surprisingly greater than the nonmutant EPO^{wt}-EPO^{wt} dimer. Mutation of a single domain of the EPO dimer resulted in an unexpectedly high *in vitro* biological activity (135 U/ml for EPO^{R103A}-EPO^{wt} and 123 U/ml for EPO^{wt}-EPO^{R103A}) as determined using the Krystal bioassay. These values were only slightly lower than nonmutant EPO^{wt}-EPO^{wt} (151 U/ml) and significantly above monomeric EPO (6.3 U/ml). Thus, inactivation of one EPO molecule in a fusion protein dimer does not, as would be expected, reduce the biological activity to a level observed with one protein molecule of EPO. Retention of enhanced biological activity

by the EPO^{R103A}-EPO^{wt} and EPO^{wt}-EPO^{R103A} mutant dimers could suggest that each of the two domains in EPO^{wt}/EPO^{wt} is capable of activating the EPO receptor on target cells such as spleen cells used in the bioassay.

Interestingly, the specific activities of EPO^{R103A}-EPO^{wt} and EPO^{wt}-EPO^{R103A} mutants were 480 U/μg and 516 U/μg, respectively, essentially one-half that of the non-mutant EPO^{wt}-EPO^{wt} dimer, indicating that the two EPO domains in the EPO^{wt}-EPO^{wt} dimer are equally active.

Northern blot analysis revealed that mRNA encoding the double EPO^{R103A}/EPO^{R103A} mutant was expressed. However, no biological activity or secreted fusion protein was detected in conditioned medium from COS1 cells transfected with EPO^{R103A}/EPO^{R103A} constructs.

EXAMPLE 9

EVALUATION OF BIOLOGICAL ACTIVITY OF RECOMBINANT POLYPEPTIDE VARIANT PROTEINS

The biological activity of the recombinant polypeptide variants containing mutations in the 5′ untranslated region, or 3′ untranslated region or both is determined using *in vitro* and *in vivo* assays.

The recombinant polypeptide variant proteins are preferably purified substantially prior to use, particularly where the protein could be employed as an *in vivo* therapeutic, although the degree of purity is not necessarily critical where the molecule is to be used *in vitro*. In one embodiment, the recombinant polypeptides can

be isolated to about 50% purity (by weight), more preferably to about 80% by weight or about 95% by weight. It is most preferred to utilize a protein which is essentially pure (e.g., about 99% by weight or to homogeneity) for *in vitro* and *in vivo* assays as well as *in vivo* therapeutics.

For example, recombinant EPO variant proteins, which can be prepared according to site-directed mutagenesis methods discussed in Examples 6 and 10, can be screened for *in vitro* and *in vivo* activity prior to use in therapeutic settings. The *in vitro* assay measures the effect of EPO variant proteins on erythropoiesis in intact mouse spleen cells assay according to the procedure of Krystal, G., *Exp. Hematol.*, 11:649-660 (1983). To screen the various recombinant EPO variant proteins for activity, for example, *in vitro* or *in vivo*, the proteins (or mixtures of the EPO proteins) can be evaluated for the extent of hematopoieses, platelet production or receptor binding. Tests to determine biological activity are well-known to those of skill in the art. For example, the biological activity of EPO can be measured as described in Sytkowski and Grodberg (U.S. Patent Nos. 5,614,184); Sytkowski (U.S. Patent No. 5,580,853); Sytkowski, U.S. patient application "Modified Polypeptides with Increased Biological Activity", filed February 3, 1998; and Powell (U.S. Patent No. 5,688,679); the teachings of which are herein incorporated by reference in their entirety.

EXAMPLE 10

POLYPEPTIDE VARIANTS PRODUCED BY ALTERING NONCODING REGIONS OF THE GENE

Typically, variants of recombinant proteins are made by deleting, adding or substituting nucleotides within the coding of the gene. However, it is also possible to make variants of recombinant proteins by altering the noncoding regions of genes, i.e., the 5' and 3' untranslated regions (UTR). Modifications in the UTR of a gene, especially in the 5' sequence as well as in the first intron, influence the regulation of translation; and, thus, the expression of proteins (Schultz, D.E., *et al.*, *J. Virol.* 70:1041-1049, 1996; Kozak, M., *J. Mol. Biol.* 235:95-110, 1994; Bettany, A.J., *et al.*, *J. Biol. Chem.* 267:16531-16537, 1992; Kozak, M., *J. Biol. Chem.* 266:19867-19870, 1991).

Alterations in the non-coding sequences of a polypeptide gene can result in different mRNA secondary structure (e.g., free energy of the loops and base pairs), translation efficiency; and subsequently, the expression, secretion and biological activity of the polypeptide. Therefore, different forms of polypeptides can be manufactured as a result of modifications in regions which flank either the 5' or 3' side of the coding region of a polypeptide.

Figure 12 is a schematic representation of changes in mRNA structure and ultimately protein structure and function that can result when an alteration(s) is made in the 5' and/or 3' UTR of the recombinant polypeptide gene.

Variations in the recombinant polypeptide can be produced as, for example, different restriction enzyme generated fragments of genomic sequences and/or specific nucleotide substitutions and mutations in the 5' and/or 3' UTR of the polypeptide coding sequence. Oligonucleotide-directed site-specific mutagenesis procedures as described herein can be employed to provide the recombinant polypeptide variant proteins.

Modifications in the noncoding regions of the polypeptide gene can affect mRNA stability, rates of translation, expression from host cells, protein processing, export from rough endoplasmic reticulum, extent and pattern of glycosylation, secretion dynamics and rates of export from the cell. For example, varied glycosylation patterns can result, which, for EPO, are of great importance for biological activity (Yamaguchi, K., *et al.*, *J. Biol. Chem.* 266:20434-20439, 1991). The resulting proteins can represent chemically, structurally and biologically distinct forms of recombinant polypeptides.

The nucleotide sequences of polypeptide variants can be confirmed by DNA sequencing using standard experimental procedures. Distinctive versions of genomic polypeptides can be produced by mutations in the 5' and 3' UTR and can be detected by Southern blotting. Likewise, different mRNAs can be identified by Northern blotting. Differences in hybridization conditions, i.e., high or low stringencies, will be an index of the diversity of the DNA and mRNA. It is possible that different genomic sequences

may require different promoters (e.g., mouse metallothionein or 3-phosphoglycerate), vectors (e.g., bovine papilloma virus), and/or host cells (e.g., CHO, BHK-21 or C127 cells) to adequately express the recombinant polypeptide. The technical methods which can be employed for the above mentioned experimental strategies are familiar to those of skill in the art. For example, detailed protocols can be found in Sambrook, *et al.*, "*Molecular Cloning: A Laboratory Manual*," (1989) and Ausubel, *et al.*, "Current Protocols in Molecular Biology, "(1995); Powell, J.S., *et al.*, *Proc. Natl. Acad. Sci. USA* 83:6465-6469, 1986; and Sytkowski and Grodberg, (U.S. Patent No. 5,614,184); Sytkowski (U.S. Patent No. 5,580,853); and Powell (U.S. Patent No. 5,688,679); the teachings of which are herein incorporated by reference in their entirety.

Mutations in the 5' and/or 3' UTR of the polypeptide gene can result in altered RNA structure, total free energy, stability and/or rates and efficiency of translation (Schultz, D.E., *et al.*, *J. Virol.* 70:1041-1049, 1996; Kozak, M., *J. Mol. Biol.* 235:95-110, 1994; Bettany, A.J., *et al.*, *J. Biol. Chem.* 267:16531-16537, 1992; Kozak, M., *J. Biol. Chem.* 266:19867-19870, 1991; Purvis, I.J., *et al.*, *Nucleic Acids Res.* 15: 7951-62, 1987). The secondary structure of mRNAs play an important role in the initiation and efficiency of translation and, thus, in protein synthesis.

Computer modeling using the PC/Gene® RNAFOLD program (IntelliGenetics, Inc.) is used to predict differences in RNA secondary structure, specifically the total free energy, following deletion in the 5' or 3' UTR of, for example, the EPO gene (Figures 13-15). The program utilizes an algorithm which calculates the energies of the secondary structure of RNA. It automatically transcribes any DNA sequence into a single stranded RNA sequence. Since the mRNA is single stranded, it can fold back upon itself due to the complementarity of bases resulting in various "loops". Energy must be released to form a base-paired or looped structure and the stability of the resulting secondary structure is determined by the amount of energy released. Therefore, if alternative structures have a free energy of formation of -50 kcal/mol and -100 kcal/mol, the latter structure is intrinsically more likely to be formed.

For example, free energy for the secondary RNA structure for nucleotides 401-624 in the 5' UTR of the EPO gene is predicted to be -161.0 kcal/mol (SEQ ID NO: 2). A 50 nucleotide deletion spanning nucleotides 501-550 results in a total free energy of -127.2 kcal/mol (SEQ ID NO: 3), whereas a 50 nucleotide deletion at nucleotides 551-600 (SEQ ID NO: 4) results in an RNA structure with -118.9 kcal/mol of free energy indicating the importance of the size of the deletion and location in ultimately defining mRNA secondary structure. Larger deletions, in different portions of the 401-624 region of the 5' UTR, yield RNA

structures with varying predicted energy states (SEQ ID NOS: 5-7). These results are summarized in Table 2.

TABLE 2: SEQUENCE VARIATION IN 5' UTR-EFFECT ON mRNA FREE ENERGY

Sequence	SEQ ID NO:	Nucleotide Length (bp)	Region of Deletion	Number of Nucleotide Deleted (bp)	Free Energy (kal/mol)
Native	2	224	---	---	-161.0
5'a	3	174	501-550	50	-127.2
5'b	4	174	551-600	50	-118.9
5'c	5	124	401-550	100	-94.1
5'd	6	74	401-550	150	-52.3
5'e	7	34	401-590	190	-11.3

Likewise, for example, the free energy for the RNA secondary structure for nucleotides 2773-2972 in the 3' UTR of the EPO gene is predicted to be -81.4 kcal/mol (SEQ ID NO: 8). A 50 nucleotide deletion spanning nucleotides 2923-2972 (SEQ ID NO: 9) results in a total free energy of -53.5 kcal/mol, whereas a 100 nucleotide deletion at nucleotides 2873-2972 (SEQ ID NO: 10) results in an RNA structure with -33.3 kcal/mol of free energy. Larger deletions, in different portions of the 2773-2973 region of the 3' UTR, yield RNA structures with varying predicted energy states (SEQ ID NOS: 11 and 12). These results are summarized in Table 3.

TABLE 3: SEQUENCE VARIATION IN 3' UTR-
EFFECT ON mRNA FREE ENERGY

Sequence	SEQ ID NO:	Nucleotide Length (bp)	Region of Deletion	Number of Nucleotide Deleted (bp)	Free Energy (kal/ mol)
Native	8	200	---	---	-81.4
3'a	9	150	2923-2972	50	-53.5
3'b	10	100	2873-2972	100	-33.3
3'c	11	50	2823-2972	150	-12.5
3'd	12	100	2801-2900	100	-36.6

The secondary structure of mRNA affects the rates of translation of the corresponding coding regions (Kikinis, Z., *et al.*, *Nucleic Acids Res. 23:* 4190-4195, 1995; Kozak, M., *Mamm. Genome 7:* 563-574, 1996; Bettany, A.J., *et al.*, *J. Biol. Chem. 267:* 16531-16537, 1992; Kozak, M., *J. Mol. Biol. 235:* 95-110, 1994). Secondary structure loops in the mRNA must be unwound to facilitate ribosome attachment and proper protein assembly (Alberts, B., *et al.*, *"Molecular Biology of the Cell"* , 3rd ed., Garland Publishing, Inc., New York, NY, pp. 223-290, 1994).

The nascent polypeptide chains can interact with chaperon proteins, for example, BiP, in unique ways which can affect the proper folding of the polypeptide chain and influence passage of the protein through the endoplasmic

reticulum thereby altering glycosylation of the resulting protein. Recent data suggest that BiP-like proteins not only bind improperly folded proteins but also may assist in the appropriate protein folding and facilitate the membrane translocation and glycosylation of secretory proteins (Knittler, M.R., *et al.*, *EMBO J.*11:1573-1581, (1992); Sanders, S.L. *et al.*, *Cell* 69:353-365, (1992)). Alterations in glycosylation patterns can influence the secretion and, in the case of EPO, drastically alter biological activity (Yamaguchi, K., *et al.*, *J. Biol. Chem.* 266:20434-20439, 1991).

The three dimensional structure of a polypeptide, for example EPO, is significantly influenced by the protein backbone and the oligosaccharide chains. Alterations in the carbohydrate composition (e.g., the number of N- or O-linked oligosaccharide residues and/or type of sugar moieties) can lead to different biological properties of the polypeptide variant proteins and, thus, varied therapeutic effects. Therefore, a difference in the 5' or 3' UTR can affect mRNA secondary structure, which in turn can influence the rate of expression and post-translational modifications such as glycosylation. The proper glycosylation of a polypeptide can be of paramount importance to proper folding and secretion of the mature product and, hence, its biological and pharmacological properties.

Indices of intrinsic structural variations in the recombinant polypeptide variant proteins can be manifested

in differences in the three-dimensional structure of the protein backbone and the extent and pattern of carbohydrate chains. For example, circular dichroism (CD) spectra and thermal stability for the resulting polypeptide variants can be performed to determine the content of alpha helix, beta sheet, beta turn and random coil for different glycoproteins. The structure of the oligosaccharide chains can be determined, for example, using enzymatic and chemical deglycosylation, gas chromatography, methylation analyses, fast-atom-bombardment mass spectrometry as well as one-and two-dimensional ^{1}H-NMR spectrometry. The methods to perform the above mentioned analyses are routine to one of ordinary skill in the art and are delineated in detail in several references including for example, Ausubel, F.M., *et al.*, *"Current Protocols in Molecular Biology"* (1995); Nimtz, M., *et al. Eur. J. Biochem. 213:* 39-56, 1993; and Nimtz, M., *et al.*, *FEBS 365:* 203-208, 1995, the teachings of which are herein incorporated by reference in their entirety.

In addition, assessment of the structural differences in the recombinant polypeptide variant proteins could be evaluated using immunoprecipitation with polypeptide-specific monoclonal antibodies and heat denaturation curves. Experimental techniques to measure these properties of a polypeptide, for example, EPO, are described in Sytkowski and Grodberg (U.S. Patent No. 5,614,184); Sytkowski (U.S. Patent No. 5,580,853); and

Powell (U.S. Patent No. 5,688,679); the teachings of which are herein incorporated by reference in their entirety.

EQUIVALENTS

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the claims.

CLAIMS

What is claimed is:

1. A fusion protein with increased biological activity comprising two or more protein molecules.

2. The fusion protein of Claim 1 having an increased rate of secretion.

3. The fusion protein of Claim 1, wherein said protein molecules comprise erythropoietin.

4. An isolated nucleic acid encoding the erythropoietin fusion protein of Claim 3 comprising SEQ ID NO: 16.

5. The erythropoietin fusion protein of Claim 3 comprising the amino acid sequence, SEQ ID NO: 17.

6. The fusion protein of Claim 1 or Claim 2, wherein said protein molecules are linked by a peptide linker.

7. The fusion protein of Claim 6, wherein said peptide linker allows the protein molecules to rotate freely relative to each other.

8. The fusion protein of Claim 6 or Claim 7, wherein said peptide linker is from about 10 amino acids to about 20 amino acids in length.

9. The fusion protein of Claim 8, wherein said peptide linker is about 15 amino acids in length.

10. The fusion protein linker of Claims 6-9, wherein said amino acids are selected from the group consisting of Glycine, Serine, Asparagine, Threonine and Alanine.

11. A nucleic acid comprising a nucleotide sequence which encodes a fusion protein with increased biological activity comprising two or more protein molecules.

12. A vector comprising said nucleic acid of Claim 11.

13. A host cell transfected with the vector of Claim 12.

14. A composition comprising a fusion protein of any one of Claims 1-10 and a pharmaceutically acceptable carrier.

15. A method for producing a fusion protein comprising culturing the cell of Claim 13 in a suitable medium to produce a fusion protein.

16. A method of treating or preventing a condition or deficiency in a mammal in need of such treatment comprising administering to the mammal a therapeutically effective amount of a fusion protein of anyone of Claims 1-10 or the fusion protein encoded by the nucleic acid of Claim 11.

17. A method of treating or preventing anemia in a mammal comprising administering to the mammal a therapeutically effective amount of said fusion protein of Claim 3 (or any one of Claims 4-10 as a dependent on Claim 3).

18. An isolated nucleic acid encoding a fusion protein comprising two or more molecules, wherein the fusion protein has altered biological activity and wherein the nucleic acid has, one or more mutations in a noncoding region of the nucleic acid.

19. The nucleic acid of Claim 18 wherein a mutation is in the 5' noncoding region.

20. The nucleic acid of Claim 19 wherein the nucleic acid encodes an erythropoietin fusion protein and comprises a nucleic acid selected from the group consisting of SEQ ID NO: 3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO: 6 and SEQ ID NO: 7.

21. A method of treating or preventing a condition or deficiency in a mammal in need of such treatment comprising administering to the mammal a therapeutically effective amount of a fusion protein encoded by the nucleic acid of any one of Claims 18-21.

22. A composition comprising a fusion protein encoded by the nucleic acid of Claims 19-20 and a pharmaceutically acceptable carrier.

23. A method of treating or preventing anemia in a mammal comprising administering to the mammal a therapeutically effective amount of a fusion protein encoded by a nucleic acid of Claim 20.

24. The nucleic acid of Claim 18 wherein a mutation is in the 3' noncoding region.

25. The nucleic acid of Claim 24 wherein the nucleic acid encodes an EPO fusion protein and comprises a nucleic acid selected from the group consisting of SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; and SEQ ID NO: 12.

26. A method of treating or preventing a condition or deficiency in a mammal in need of such treatment comprising administering to the mammal a

therapeutically effective amount of a fusion protein encoded by the nucleic acid of Claim 24 and Claim 25.

27. A composition comprising a fusion protein of Claim 24 or Claim 25 and a pharmaceutically acceptable carrier.

28. A method of treating or preventing anemia in a mammal comprising administering to the mammal a therapeutically effective amount of a fusion protein encoded by a nucleic acid of Claim 25.

29. The nucleic acid of Claim 18 wherein a mutation is in both the 5' and 3' noncoding regions.

30. The nucleic acid of Claim 29 wherein the nucleic acid encodes an EPO fusion protein and comprises a nucleic acid selected from the group consisting of SEQ ID NO:3; SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; and SEQ ID NO: 12.

31. A method of treating or preventing a condition or deficiency in a mammal in need of such treatment comprising administering to the mammal a therapeutically effective amount of a fusion protein encoded by the nucleic acid of Claim 29.

32. A composition comprising a fusion protein of Claim 29 and a pharmaceutically acceptable carrier.

33. A method of treating or preventing anemia in a mammal comprising administering to the mammal a therapeutically effective amount of a fusion protein encoded by a nucleic acid of Claim 30.

34. A fusion protein with increased biological activity comprising two or more protein molecules wherein at least one protein molecule has one or more mutations.

35. The fusion protein of Claim 34 wherein said protein molecules comprise erythropoietin.

36. The erythropoietin fusion protein of Claim 35 wherein at least one of the erythropoietin molecules has a mutation wherein the arginine residue at position 103 is replaced with alanine.

37. A nucleic acid encoding a fusion protein of Claim 34.

38. A nucleic acid encoding a fusion protein of Claim 36.

39. A method of treating or preventing anemia in a mammal comprising administering to the mammal a therapeutically effective amount of the erythropoietin fusion protein of Claim 36.

40. A composition comprising a fusion protein of Claim 36 and a pharmaceutically acceptable carrier.

41. The fusion protein of any one of Claims 1-10 or the fusion encoded by the nucleic acid of any one of Claims 11, 18-20, 24-25 and 29-30 for use in therapy or prophylaxis, for example in treating or preventing a condition or deficiency in a mammal in need of such treatment.

42. Use of the fusion protein of any one of Claims 1-10 or the fusion protein encoded by the nucleic acid of any one of Claims 11, 18-20, 24-25 and 29-30 for the manufacture of a medicament for use in therapy or prophylaxis, for example in treating or preventing a condition or deficiency in a mammal in need of such treatment.

43. A process for the production of a therapeutic or prophylactic composition (for example, for treating or preventing a condition or deficiency in a mammal in need of such treatment) characterized in the use, as an essential constituent of said composition, of the fusion protein or any one of Claims 1-10 or the fusion protein encoded by the nucleic acid of any one of Claims 11, 18-20, 24-25 and 29-30.

44. The fusion protein of Claim 41, use of Claim 42 or process of Claim 43 wherein the fusion protein comprises erythropoietin and the therapy or prophylaxis is the treatment or prevention of anaemia.

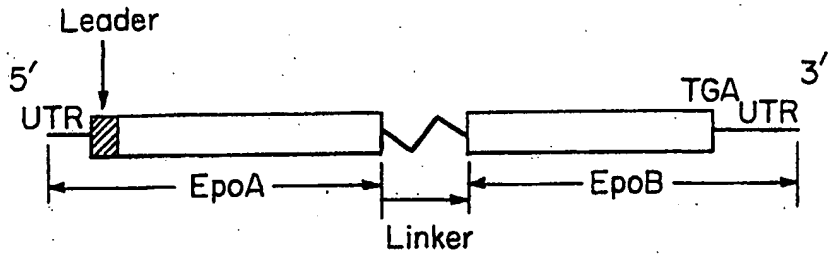


FIG. 1

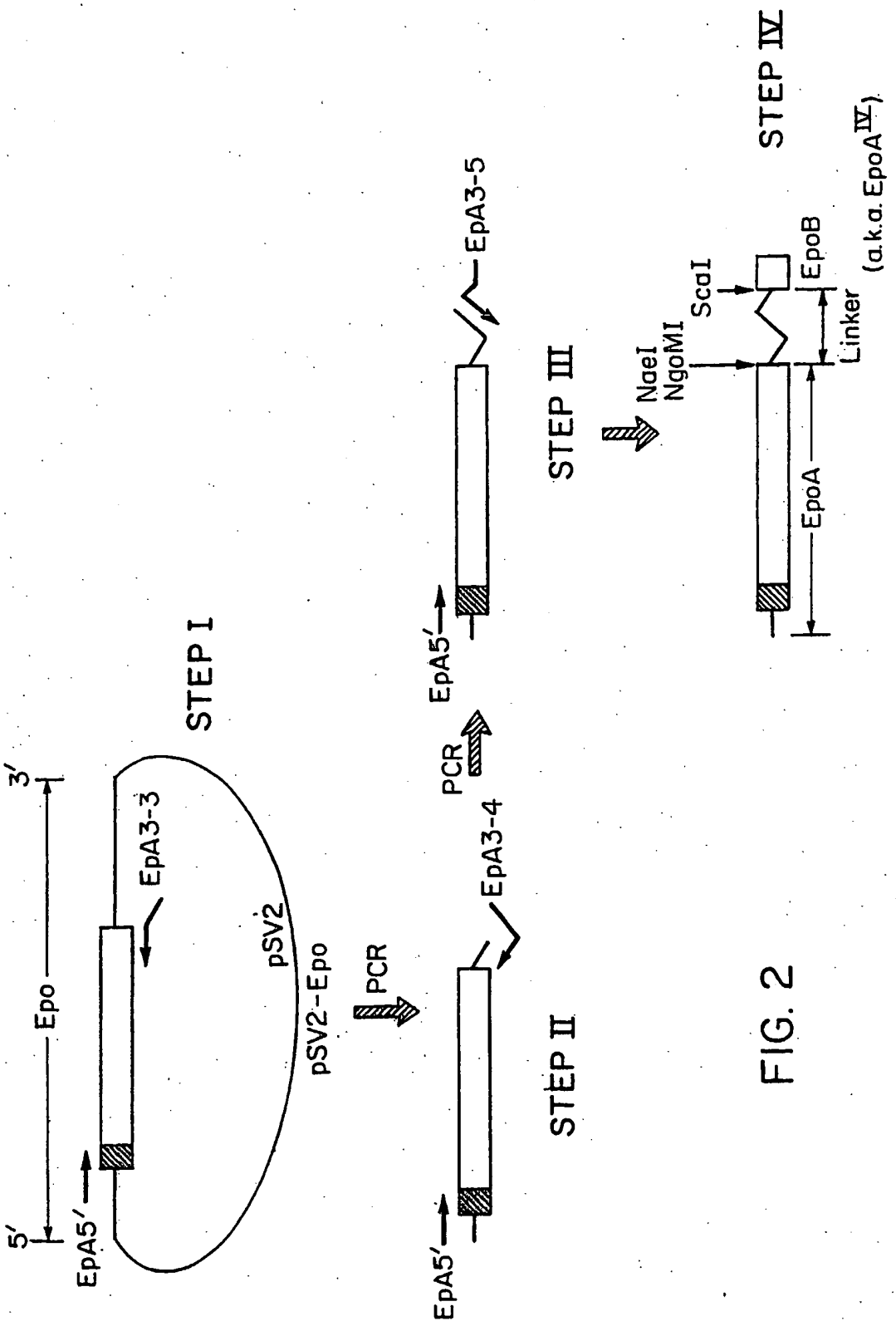

STEP I
3'
5'
Epo
EpA3-3
EpA5'
pSV2
pSV2-Epo
PCR
EpA3-4
EpA5'
STEP II
PCR
EpA5'
EpA3-5
STEP III
NaeI
NgoMI
ScaI
EpoB
Linker
EpoA
(a.k.a. EpoA^IV)
STEP IV

FIG. 2

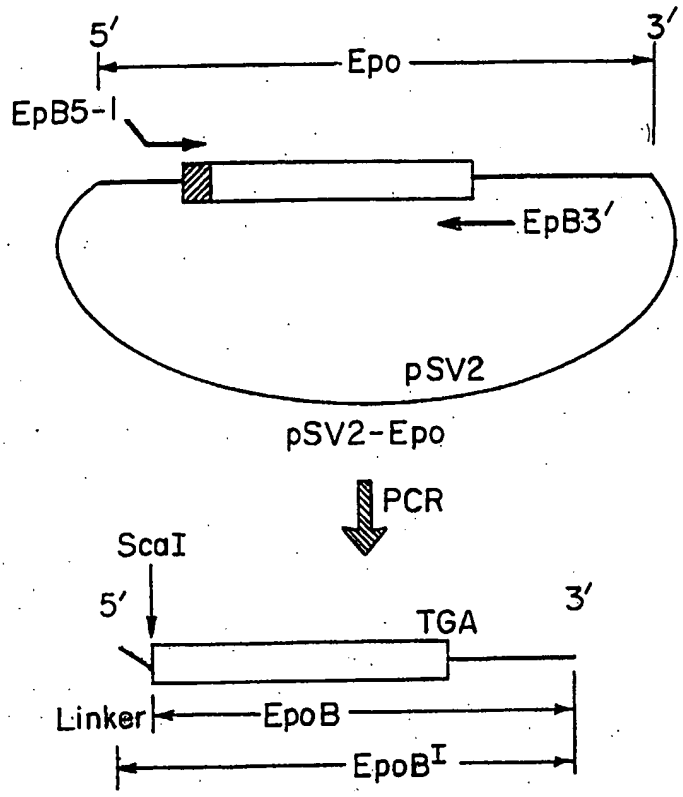

5′
3′
Epo
EpB5-I
EpB3′
pSV2
pSV2-Epo
PCR
ScaI
5′
TGA
3′
Linker
EpoB
EpoB^I

FIG. 3

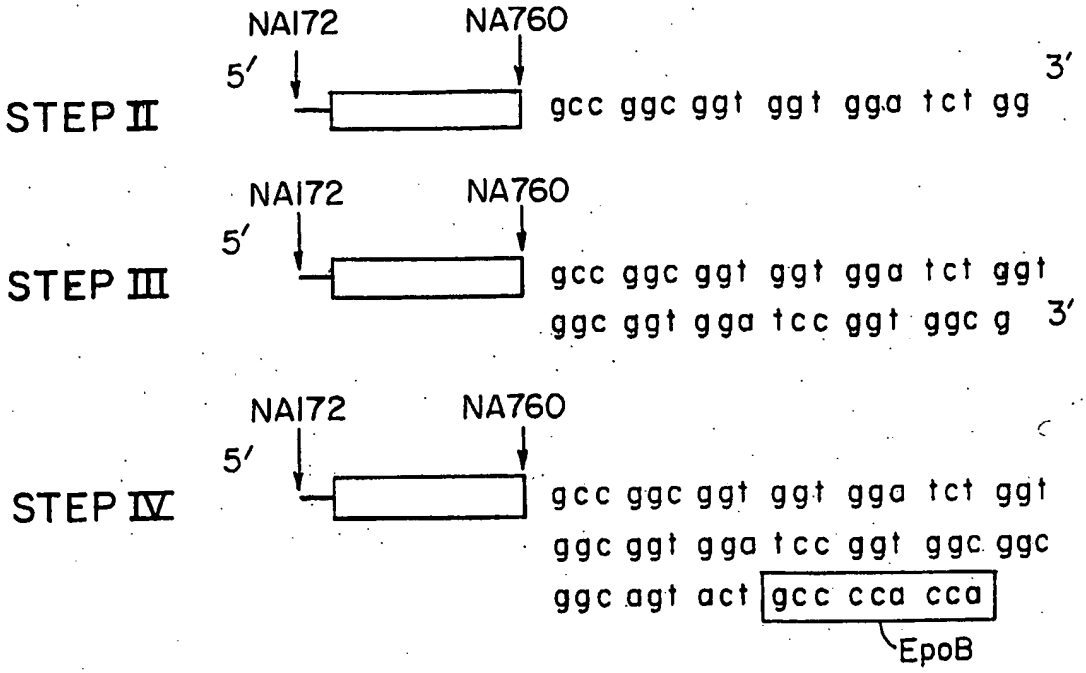

NAI72
NA760
5′
3′
STEP II
gcc ggc ggt ggt gga tct gg
NAI72
NA760
5′
STEP III
gcc ggc ggt ggt gga tct ggt
ggc ggt gga tcc ggt ggc g 3′
NAI72
NA760
5′
STEP IV
gcc ggc ggt ggt gga tct ggt
ggc ggt gga tcc ggt ggc ggc
ggc agt act gcc cca cca
EpoB

FIG. 4

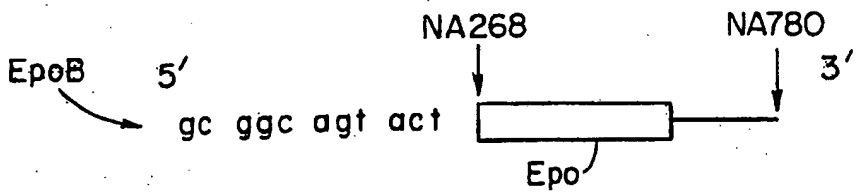

NA268
NA780
EpoB
5′
3′
gc ggc agt act
Epo

FIG. 5

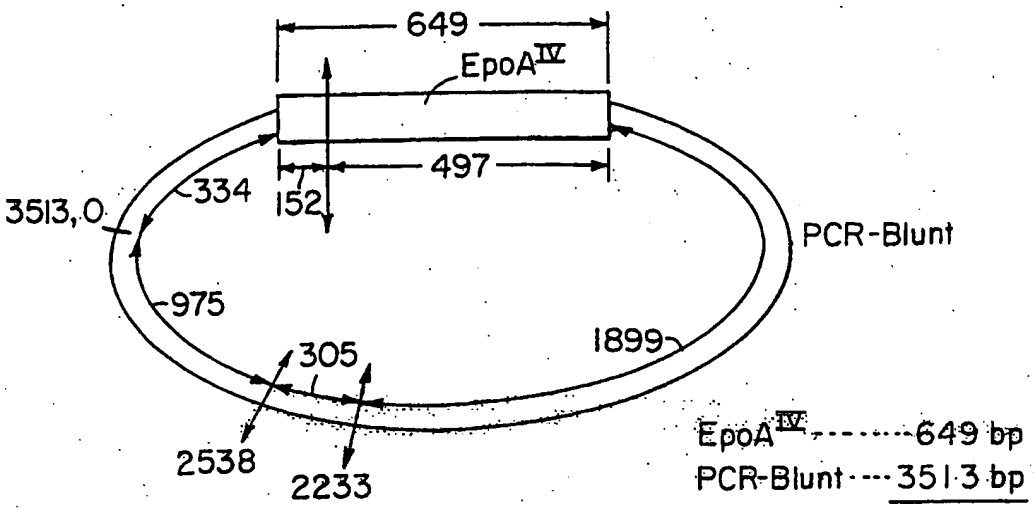


FIG. 6 BglI restriction digests: ① EpoA^{IV} Forward Direction ⊕
2396bp, 1461bp, 305bp

② EpoA^{IV} Reverse Direction ⊖
2051bp, 1806bp, 305bp

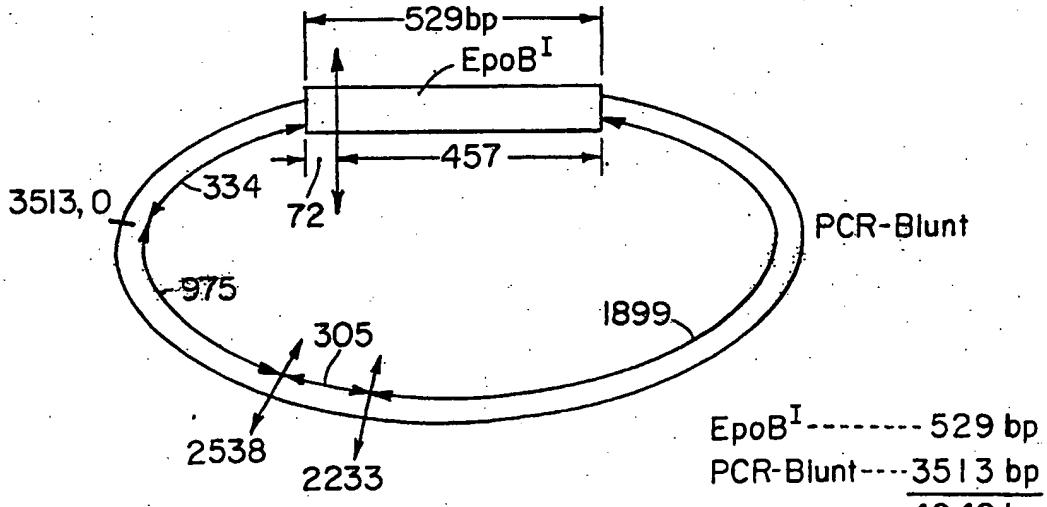


FIG. 7 BglI restriction digests: ① EpoB^{I} Forward Direction ⊕
2356bp, 1381bp, 305bp

② EpoB^{I} Reverse Direction ⊖
1971bp, 1766bp, 305bp

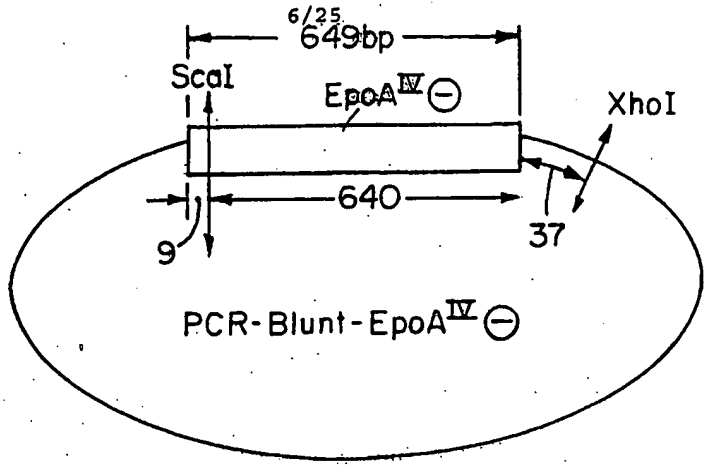


FIG. 8 ScaI and XhoI restriction digests: 3485bp, 6776bp

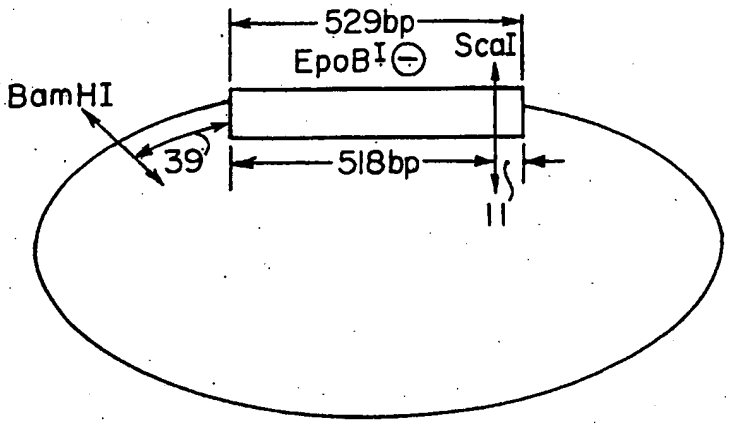


FIG. 9 BamHI and ScaI restriction digests: 3485bp, 5576bp

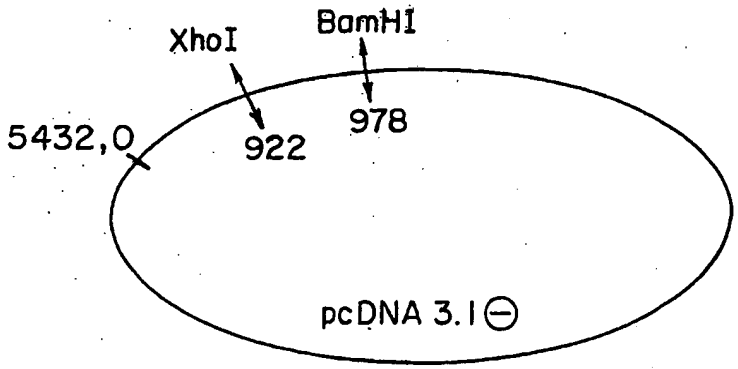
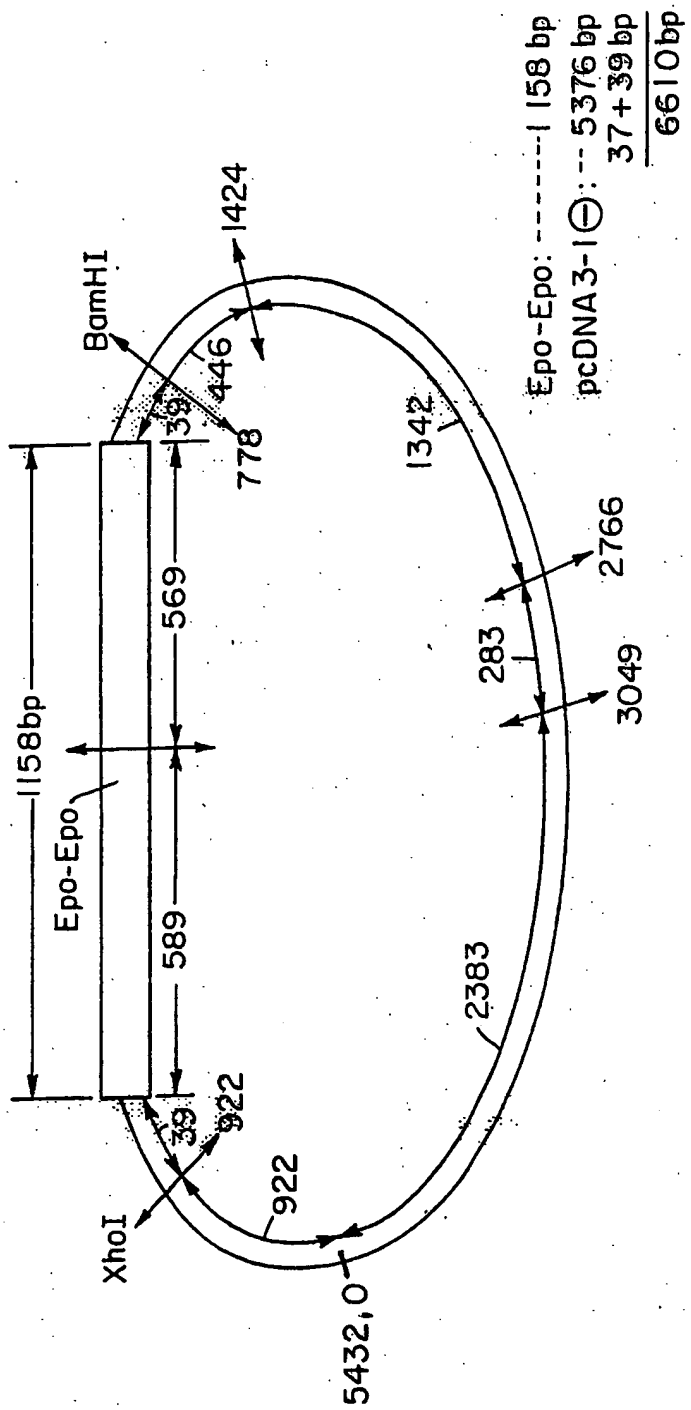


FIG. 10 XhoI and BamHI restriction digest: 5376bp, 56bp



NgoMI restriction digest of pcDNA3-1-Epo-Epo: 3931bp, 1342bp, 1054bp, 2836bp

FIG. 11

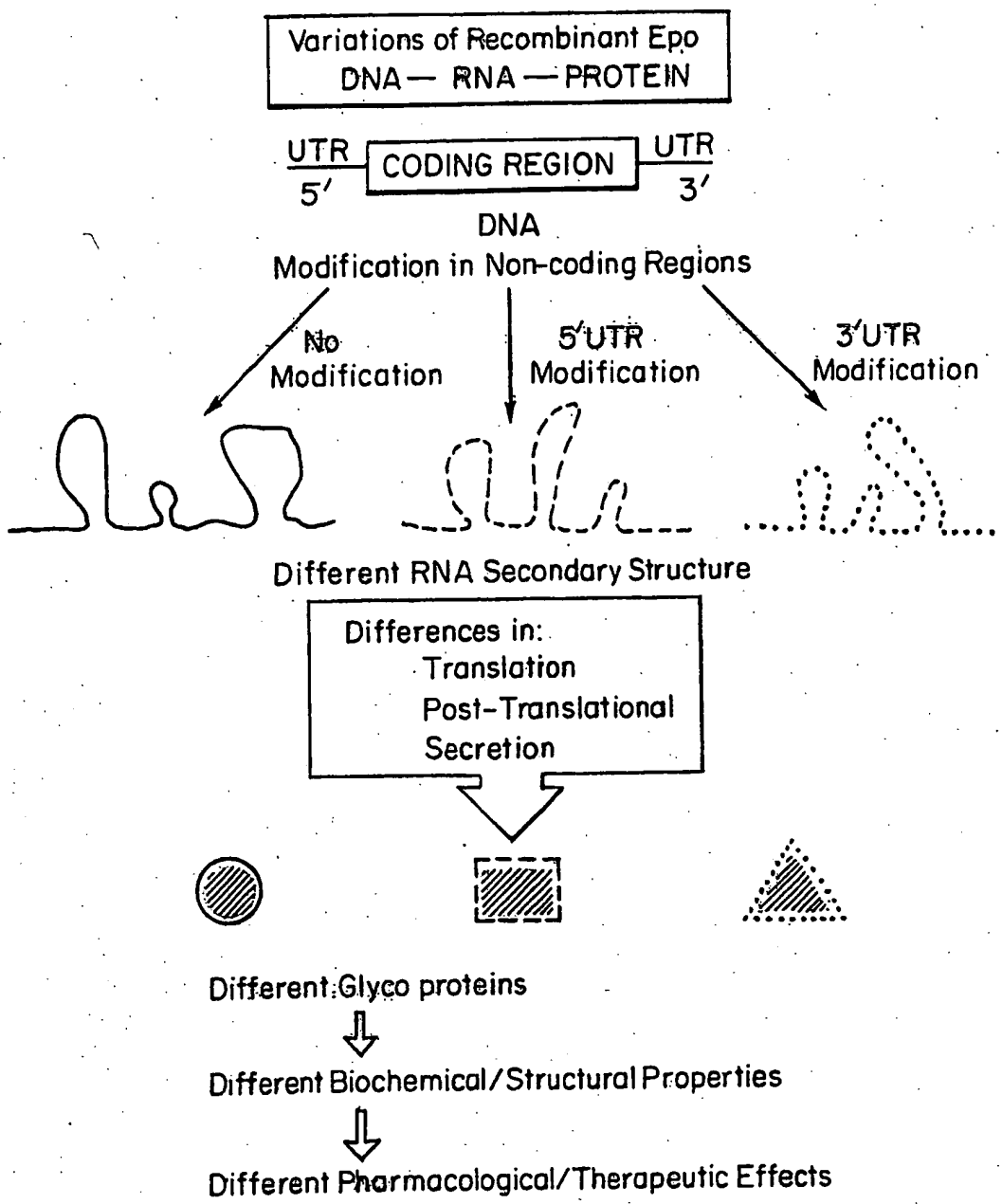

Variations of Recombinant Epo
DNA — RNA — PROTEIN
UTR
5′
CODING REGION
UTR
3′
DNA
Modification in Non-coding Regions
No Modification
5′UTR Modification
3′UTR Modification
Different RNA Secondary Structure
Differences in:
Translation
Post-Translational
Secretion
Different Glyco proteins
Different Biochemical/Structural Properties
Different Pharmacological/Therapeutic Effects

FIG. 12

```
aagcttctgg gcttccagac ccagctactt tgcggaactc agcaacccag gcatctctga      60
gtctccgccc aagaccggga tgccccccag gggaggtgtc cgggagccca gcctttccca     120
gatagcacgc tccgccagtc ccaagggtgc gcaaccggct gcactccccc cccgcgaccc     180
agggcccggg agcagccccc atgacccaca cgcacgtctg cagcagcccc gctcacgccc     240
cggcgagcct caacccaggc gtcctgcccc tgctctgacc ccgggtggcc cctacccctg     300
gcgacccctc acgcacacag cctctccccc acccccaccc gcgcacgcac acatgcagat     360
aacagccccg acccccggcc agagccgcag agtccctggg ccaccccggc cgctcgctgc     420
gctgcgccgc accgcgctgt cctcccggag ccggaccggg gccaccgcgc ccgctctgct     480
ccgacaccgc gccccctgga cagccgccct ctcctctagg cccgtggggc tggccctgca     540
ccgccgagct tcccgggatg agggcccccg gtgtggtcac ccggcgcgcc ccaggtcgct     600
gagggacccc ggccaggcgc ggagatgggg gtgcacggtg agtactcgcg ggctgggcgc     660
tcccgccgcc cgggtccctg tttgagcggg gatttagcgc cccggctatt ggccaggagg     720
tggctgggtt caaggaccgg cgacttgtca aggaccccgg aaggggggagg ggggtggggc     780
agcctccacg tgccagcggg gacttggggg agtccttggg gatggcaaaa acctgacctg     840
tgaaggggac acagtttggg ggttgagggg aagaaggttt gggggttctg ctgtgccagt     900
ggagaggaag ctgataagct gataacctgg gcgctggagc caccacttat ctgccagagg     960
ggaagcctct gtcacaccag gattgaagtt tggccggaga agtggatgct ggtagctggg    1020
ggtggggtgt gcacacggca gcaggattga atgaaggcca gggaggcagc acctgagtgc    1080
ttgcatggtt ggggacagga aggacgagct ggggcagaga cgtggggatg aaggaagctg    1140
tccttccaca gccacccttc tccctccccg cctgactctc agcctggcta tctgttctag    1200
aatgtcctgc ctggctgtgg cttctcctgt ccctgctgtc gctccctctg ggcctcccag    1260
tcctgggcgc cccaccacgc ctcatctgtg acagccgagt cctggagagg tacctcttgg    1320
```

Figure 13A

```
aggccaagga ggccgagaat atcacggtga gacccttcc ccagcacatt ccacagaact    1380
cacgctcagg gcttcaggga actcctccca gatccaggaa cctggcactt ggtttggggt    1440
ggagttggga agctagacac tgccccccta cataagaata agtctggtgg ccccaaacca    1500
tacctggaaa ctaggcaagg agcaaagcca gcagatccta cggcctgtgg gccagggcca    1560
gagccttcag ggacccttga ctccccgggc tgtgtgcatt tcagacgggc tgtgctgaac    1620
actgcagctt gaatgagaat atcactgtcc cagacaccaa agttaatttc tatgcctgga    1680
agaggatgga ggtgagttcc ttttttttt ttttccttt ctttggaga atctcatttg    1740
cgagcctgat tttggatgaa agggagaatg atcgggggaa aggtaaaatg gagcagcaga    1800
gatgaggctg cctgggcgca gaggctcacg tctataatcc caggctgaga tggccgagat    1860
gggagaattg cttgagccct ggagtttcag accaacctag gcagcatagt gagatccccc    1920
atctctacaa acatttaaaa aaattagtca ggtgaagtgg tgcatggtgg tagtcccaga    1980
tatttggaag gctgaggcgg gaggatcgct tgagcccagg aatttgaggc tgcagtgagc    2040
tgtgatcaca ccactgcact ccagcctcag tgacagagtg aggccctgtc tcaaaaaaga    2100
aaagaaaaaa gaaaaataat gagggctgta tggaatacat tcattattca ttcactcact    2160
cactcactca ttcattcatt cattcattca acaagtctta ttgcatacct tctgtttgct    2220
cagcttggtg cttggggctg ctgaggggca ggagggagag ggtgacatgg gtcagctgac    2280
tcccagagtc cactccctgt aggtcgggca gcaggccgta gaagtctggc agggcctggc    2340
cctgctgtcg gaagctgtcc tgcggggcca ggccctgttg gtcaactctt cccagccgtg    2400
ggagcccctg cagctgcatg tggataaagc cgtcagtggc cttcgcagcc tcaccactct    2460
gcttcgggct ctgggagccc aggtgagtag gagcggacac ttctgcttgc cctttctgta    2520
agaaggggag aagggtcttg ctaaggagta caggaactgt ccgtattcct tccctttctg    2580
tggcactgca gcgacctcct gttttctcct tggcagaagg aagccatctc ccctccagat    2640
gcggcctcag ctgctccact ccgaacaatc actgctgaca ctttccgcaa actcttccga    2700
gtctactcca atttcctccg gggaaagctg aagctgtaca caggggaggc ctgcaggaca    2760
ggggacagat gaccaggtgt gtccacctgg gcatatccac cacctccctc accaacattg    2820
cttgtgccac accctccccc gccactcctg aaccccgtcg aggggctctc agctcagcgc    2880
cagcctgtcc catggacact ccagtgccag caatgacatc tcaggggcca gaggaactgt    2940
```

Figure 13B

```
ccagagagca actctgagat ctaaggatgt cacagggcca acttgagggc ccagagcagg    3000
aagcattcag agagcagctt taaactcagg gacagagcca tgctgggaag acgcctgagc    3060
tcactcggca ccctgcaaaa tttgatgcca ggacacgctt tggaggcgat ttacctgttt    3120
tcgcacctac catcagggac aggatgacct ggagaactta ggtggcaagc tgtgacttct    3180
ccaggtctca cgggcatggg cactcccttg gtggcaagag ccccccttgac accggggtgg    3240
tgggaaccat gaagacagga tgggggctgg cctctggctc tcatggggtc caagttttgt    3300
gtattcttca acctcattga caagaactga aaccaccaat atgactcttg gcttttctgt    3360
tttctgggaa cctccaaatc ccctggctct gtcccactcc tggcagcagt gcagcaggtc    3420
caggtccggg aaatgagggg tggagggggc tgggccctac gtgctgtctc acacagcctg    3480
tctgacctct cgacctaccg gcctaggcca caagctctgc ctacgctggt caataaggtg    3540
tctccattca aggcctcacc gcagtaaggc agctgccaac cctgcccagg gcaaggctgc    3600
ag                                                                   3602
```

Figure 13C

```
ccaccccggc cgctcgctgc gctgcgccgc accgcgctgt cctcccggag ccggaccggg     60
gccaccgcgc ccgctctgct ccgacaccgc gccccctgga cagccgccct ctcctctagg    120
cccgtggggc tggccctgca ccgccgagct tcccgggatg agggcccccg gtgtggtcac    180
ccggcgcgcc ccaggtcgct gagggacccc ggccaggcgc ggag                     224
```

Figure 14A

```
ccaccccggc cgctcgctgc gctgcgccgc accgcgctgt cctcccggag ccggaccggg      60
gccaccgcgc ccgctctgct ccgacaccgc gccccctgga tcccgggatg agggcccccg     120
gtgtggtcac ccggcgcgcc ccaggtcgct gagggacccc ggccaggcgc ggag           174
```

Figure 14B

```
ccaccccggc cgctcgctgc gctgcgccgc accgcgctgt cctcccggag ccggaccggg      60
gccaccgcgc ccgctctgct ccgacaccgc gcccccctgga cagccgccct ctcctctagg     120
cccgtggggc tggccctgca ccgccgagct gagggacccc ggccaggcgc ggag            174
```

Figure 14C

```
cagccgccct ctcctctagg cccgtggggc tggccctgca ccgccgagct tcccgggatg      60
agggcccccg gtgtggtcac ccggcgcgcc ccaggtcgct gagggacccc ggccaggcgc     120
ggag                                                                  124
```

Figure 14D

tcccgggatg agggcccccg gtgtggtcac ccggcgcgcc ccaggtcgct gagggacccc 60
ggccaggcgc ggag 74

Figure 14E

ccaggtcgct gagggacccc ggccaggcgc ggag 34

Figure 14F

```
ccaggtgtgt ccacctgggc atatccacca cctccctcac caacattgct tgtgccacac      60
cctccccccgc cactcctgaa ccccgtcgag gggctctcag ctcagcgcca gcctgtccca     120
tggacactcc agtgccagca atgacatctc aggggccaga ggaactgtcc agagagcaac     180
tctgagatct aaggatgtca                                                 200
```

Figure 15A

```
ccaggtgtgt ccacctgggc atatccacca cctccctcac caacattgct tgtgccacac      60
cctccccccgc cactcctgaa ccccgtcgag gggctctcag ctcagcgcca gcctgtccca     120
tggacactcc agtgccagca atgacatctc                                      150
```

Figure 15B

```
ccaggtgtgt ccacctgggc atatccacca cctccctcac caacattgct tgtgccacac      60
cctcccccgc cactcctgaa ccccgtcgag gggctctcag                           100
```

Figure 15C

```
ccaggtgtgt ccacctgggc atatccacca cctccctcac caacattgct          50
```

Figure 15D

```
ccaggtgtgt ccacctgggc atatccaccc agtgccagca atgacatctc aggggccaga   60
ggaactgtcc agagagcaac tctgagatct aaggatgtca                        100
```

Figure 15E

```
aggcgcggag atg ggg gtg cac gaa tgt cct gcc tgg ctg tgg ctt ctc        49
           Met Gly Val His Glu Cys Pro Ala Trp Leu Trp Leu Leu
            1               5                  10

ctg tcc ctg ctg tcg ctc cct ctg ggc ctc cca gtc ctg ggc gcc cca        97
Leu Ser Leu Leu Ser Leu Pro Leu Gly Leu Pro Val Leu Gly Ala Pro
     15                  20                  25

cca cgc ctc atc tgt gac agc cga gtc ctg gag agg tac ctc ttg gag       145
Pro Arg Leu Ile Cys Asp Ser Arg Val Leu Glu Arg Tyr Leu Leu Glu
 30                  35                  40                  45

gcc aag gag gcc gag aat atc acg acg ggc tgt gct gaa cac tgc agc       193
Ala Lys Glu Ala Glu Asn Ile Thr Thr Gly Cys Ala Glu His Cys Ser
                 50                  55                  60

ttg aat gag aat atc act gtc cca gac acc aaa gtt aat ttc tat gcc       241
Leu Asn Glu Asn Ile Thr Val Pro Asp Thr Lys Val Asn Phe Tyr Ala
            65                  70                  75

tgg aag agg atg gag gtc ggg cag cag gcc gta gaa gtc tgg cag ggc       289
Trp Lys Arg Met Glu Val Gly Gln Gln Ala Val Glu Val Trp Gln Gly
         80                  85                  90

ctg gcc ctg ctg tcg gaa gct gtc ctg cgg ggc cag gcc ctg ttg gtc       337
Leu Ala Leu Leu Ser Glu Ala Val Leu Arg Gly Gln Ala Leu Leu Val
     95                  100                 105
```

Figure 16A

```
aac tct tcc cag ccg tgg gag ccc ctg cag ctg cat gtg gat aaa gcc      385
Asn Ser Ser Gln Pro Trp Glu Pro Leu Gln Leu His Val Asp Lys Ala
110                 115                 120                 125

gtc agt ggc ctt cgc agc ctc acc act ctg ctt cgg gct ctg gga gcc      433
Val Ser Gly Leu Arg Ser Leu Thr Thr Leu Leu Arg Ala Leu Gly Ala
                130                 135                 140

cag aag gaa gcc atc tcc cct cca gat gcg gcc tca gct gct cca ctc      481
Gln Lys Glu Ala Ile Ser Pro Pro Asp Ala Ala Ser Ala Ala Pro Leu
            145                 150                 155

cga aca atc act gct gac act ttc cgc aaa ctc ttc cga gtc tac tcc      529
Arg Thr Ile Thr Ala Asp Thr Phe Arg Lys Leu Phe Arg Val Tyr Ser
        160                 165                 170

aat ttc ctc cgg gga aag ctg aag ctg tac aca ggg gag gcc tgc agg      577
Asn Phe Leu Arg Gly Lys Leu Lys Leu Tyr Thr Gly Glu Ala Cys Arg
    175                 180                 185

aca ggg gac aga gcc ggc ggt ggt gga tct ggt ggc ggt gga tcc ggt      625
Thr Gly Asp Arg Ala Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly
190                 195                 200                 205

ggc ggc ggc agt act gcc cca cca cgc ctc atc tgt gac agc cga gtc      673
Gly Gly Gly Ser Thr Ala Pro Pro Arg Leu Ile Cys Asp Ser Arg Val
                210                 215                 220

ctg gag agg tac ctc ttg gag gcc aag gag gcc gag aat atc acg acg      721
Leu Glu Arg Tyr Leu Leu Glu Ala Lys Glu Ala Glu Asn Ile Thr Thr
            225                 230                 235
```

Figure 16B

```
ggc tgt gct gaa cac tgc agc ttg aat gag aat atc act gtc cca gac      769
Gly Cys Ala Glu His Cys Ser Leu Asn Glu Asn Ile Thr Val Pro Asp
        240                 245                 250

acc aaa gtt aat ttc tat gcc tgg aag agg atg gag gtc ggg cag cag      817
Thr Lys Val Asn Phe Tyr Ala Trp Lys Arg Met Glu Val Gly Gln Gln
    255                 260                 265

gcc gta gaa gtc tgg cag ggc ctg gcc ctg ctg tcg gaa gct gtc ctg      865
Ala Val Glu Val Trp Gln Gly Leu Ala Leu Leu Ser Glu Ala Val Leu
270                 275                 280                 285

cgg ggc cag gcc ctg ttg gtc aac tct tcc cag ccg tgg gag ccc ctg      913
Arg Gly Gln Ala Leu Leu Val Asn Ser Ser Gln Pro Trp Glu Pro Leu
                290                 295                 300

cag ctg cat gtg gat aaa gcc gtc agt ggc ctt cgc agc ctc acc act      961
Gln Leu His Val Asp Lys Ala Val Ser Gly Leu Arg Ser Leu Thr Thr
            305                 310                 315

ctg ctt cgg gct ctg gga gcc cag aag gaa gcc atc tcc cct cca gat     1009
Leu Leu Arg Ala Leu Gly Ala Gln Lys Glu Ala Ile Ser Pro Pro Asp
        320                 325                 330

gcg gcc tca gct gct cca ctc cga aca atc act gct gac act ttc cgc     1057
Ala Ala Ser Ala Ala Pro Leu Arg Thr Ile Thr Ala Asp Thr Phe Arg
    335                 340                 345

aaa ctc ttc cga gtc tac tcc aat ttc ctc cgg gga aag ctg aag ctg     1105
Lys Leu Phe Arg Val Tyr Ser Asn Phe Leu Arg Gly Lys Leu Lys Leu
350                 355                 360                 365

tac aca ggg gag gcc tgc agg aca ggg gac aga tgaccaggtg tgtccacctg   1158
Tyr Thr Gly Glu Ala Cys Arg Thr Gly Asp Arg
                370                 375
```

Figure 16C

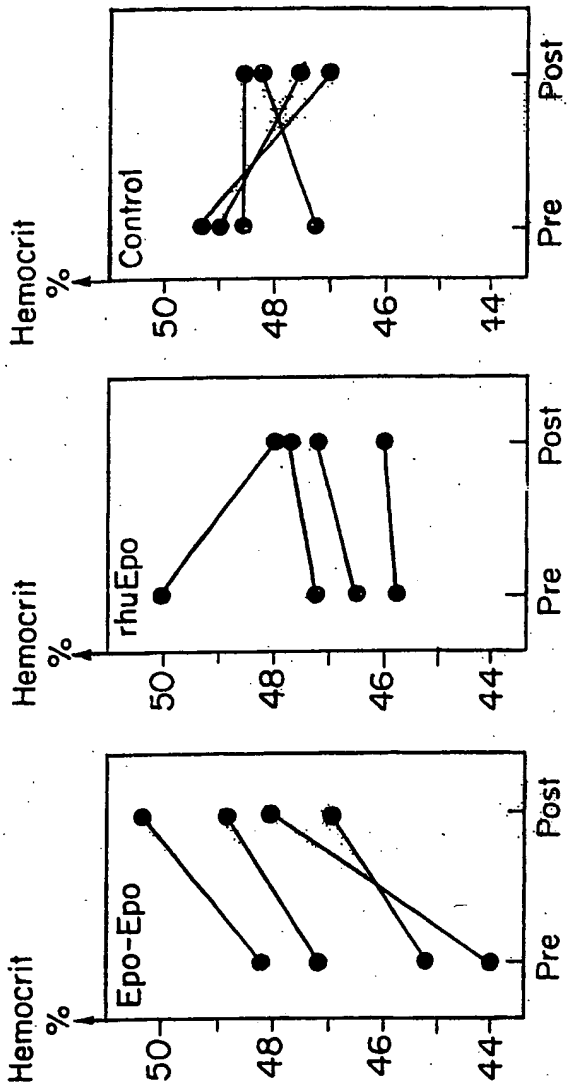


FIG. 17A

FIG. 17B

FIG. 17C

INTERNATIONAL SEARCH REPORT

i. national Application No
PCT/US 98/13944

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/62 C07K19/00 C07K14/505 C12N15/85 C12N5/10
A61K38/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 25746 A (NEW ENGLAND DEACONESS HOSPITAL) 28 September 1995 see page 3, line 2 - page 5, line 29 see page 13, line 27 - page 19, line 10; examples 2,3 ---	1,3,14, 16,17, 41-44
X Y	WO 95 33057 A (MENARINI RICERCHE SUD S.P.A.) 7 December 1995 see page 1, line 4 - line 10 see page 4, line 8 - line 21 see page 6, line 6 - page 13, line 8 --- -/--	1,3,6, 8-17, 41-44 6-10,18, 21-24, 26-28, 34-40

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 November 1998	04/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Montero Lopez, B

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

rnational Application No
PCT/US 98/13944

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 622 459 A (SANWA KAGAKU KENKYUSHO CO., LTD.) 2 November 1994 see page 4, line 38 - page 5, line 35	1,2,6,8, 11-17, 34,37, 41-43
X	LUNN E D ET AL: "Erythropoietin dimers with enhanced in vivo activity in mice." THIRTY-EIGHTH ANNUAL MEETING OF THE AMERICAN SOCIETY OF HEMATOLOGY, ORLANDO, FLORIDA, USA, DECEMBER 6-10, 1996. BLOOD 88 (10 SUPPL. 1 PART 1-2). 1996. 543A. ISSN: 0006-4971, XP002084634 see abstract 2161	1,3,14, 16,17, 41-44
Y	WO 92 06116 A (ORTHO PHARMACEUTICAL CORPORATION) 16 April 1992 see page 3, line 3 - line 16 see page 7, line 10 - line 17 see page 8, line 8 - line 24 see page 8, line 34 - page 9, line 35; examples 1,3	6-10
Y	MCGARY E C ET AL: "Post-transcriptional regulation of erythropoietin mRNA stability by erythropoietin mRNA-binding protein." JOURNAL OF BIOLOGICAL CHEMISTRY, (1997 MAR 28) 272 (13) 8628-34. JOURNAL CODE: HIV. ISSN: 0021-9258., XP002084635 United States see abstract see page 8630, left-hand column, paragraph 3 see page 8632, right-hand column, paragraph 3 - page 8634, right-hand column, paragraph 2	18, 21-24, 26-28
Y	WO 94 02611 A (NEW ENGLAND DEACONESS HOSPITAL) 3 February 1994 see page 3, line 13 - line 29 see page 6, line 3 - line 10 & US 5 614 184 A cited in the application	34-40
A	WO 96 19573 A (CANGENE CORPORATION) 27 June 1996 see example 2	4,5,20, 25,30

-/--

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national Application No
PCT/US 98/13944

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP 0 816 510 A (TORAY INDUSTRIES, INC.) 7 January 1998 see page 4, line 35 - page 5, line 25 see page 6, line 24 - line 42 see page 7, line 31 - line 45 & WO 97 23639 A (TORAY INDUSTRIES, INC.) 3 July 1997	1,3,6,8, 11-17, 42-44
P,X	SYTKOWSKI A J ET AL: "Human erythropoietin dimers with markedly enhanced in vivo activity." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1998 FEB 3) 95 (3) 1184-8. JOURNAL CODE: PV3. ISSN: 0027-8424., XP002084636 United States see abstract see page 1185, left-hand column, paragraph 4 - page 1188, left-hand column, paragraph 1	1,3,14, 16,17, 41-44
P,X	SYTKOWSKI A J ET AL: "An Epo - Epo fusion protein with enhanced potency and efficacy in vivo." 39TH ANNUAL MEETING OF THE AMERICAN SOCIETY OF HEMATOLOGY, SAN DIEGO, CALIFORNIA, USA, DECEMBER 5-9, 1997. BLOOD 90 (10 SUPPL. 1 PART 1). 1997. 57A. ISSN: 0006-4971, XP002084637 see abstract 244	1,3-6,8, 9,11-17, 34-44

2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 98/ 13944

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 16, 17, 21, 23, 28, 31, 33, and 39 is(are) directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

I. national Application No
PCT/US 98/13944

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9525746 A	28-09-1995	US 5580853 A	03-12-1996
		EP 0751959 A	08-01-1997
		US 5747446 A	05-05-1998
WO 9533057 A	07-12-1995	IT FI940106 A	27-11-1995
		AU 2616895 A	21-12-1995
		CA 2191306 A	07-12-1995
		CN 1150458 A	21-05-1997
		EP 0760858 A	12-03-1997
		JP 10500579 T	20-01-1998
EP 622459 A	02-11-1994	JP 6306100 A	01-11-1994
WO 9206116 A	16-04-1992	AU 1157695 A	13-04-1995
		AU 8735991 A	28-04-1992
		CA 2069746 A	29-03-1992
		EP 0503050 A	16-09-1992
		JP 5502463 T	28-04-1993
		PT 99107 A	31-08-1992
		ZA 9107766 A	29-03-1993
WO 9402611 A	03-02-1994	EP 0667903 A	23-08-1995
		US 5614184 A	25-03-1997
WO 9619573 A	27-06-1996	AU 692968 B	18-06-1998
		AU 4169296 A	10-07-1996
		CA 2208013 A	27-06-1996
		EP 0800577 A	15-10-1997
EP 816510 A	07-01-1998	CA 2213512 A	03-07-1997
		WO 9723639 A	03-07-1997